**Appendix S1**

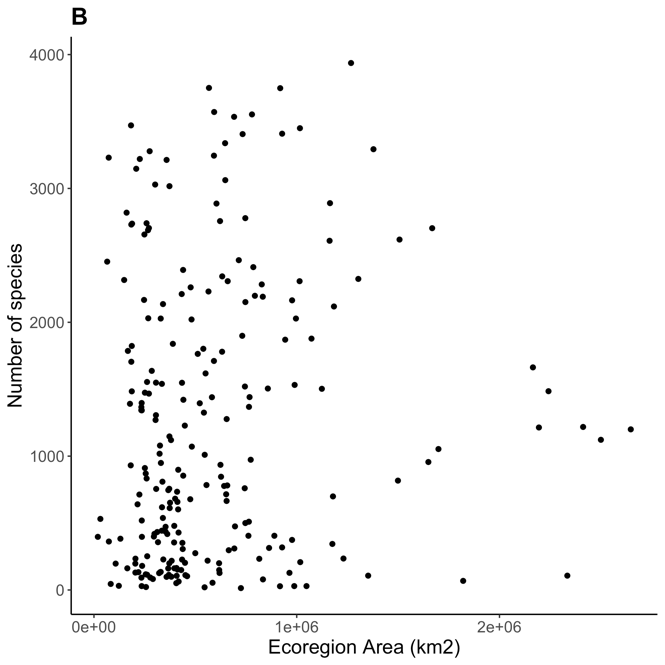
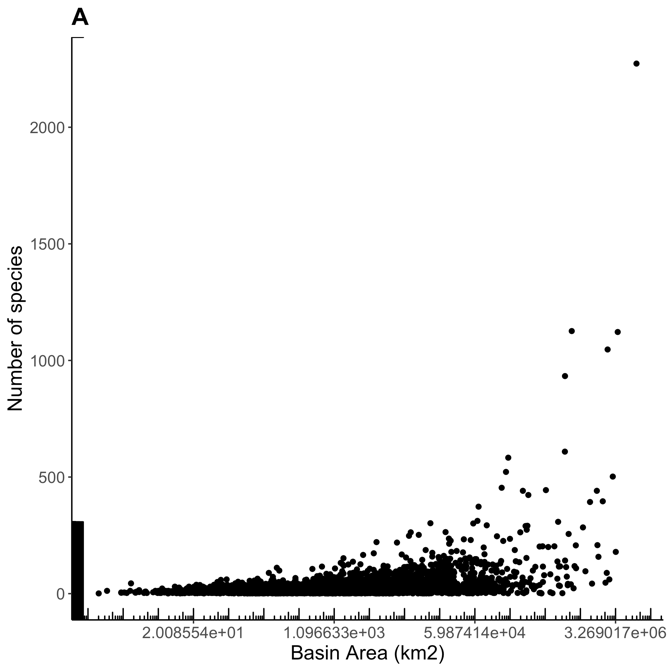
**GAPeDNA: Assessing and mapping global species gaps in genetic databases for eDNA metabarcoding**

**Table S1**: Primer pairs used for the present study, with details on the marker, primer name, primer sequences, and reference paper.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Marker | Primer name | Forward (5'-3') | Reverse (5'-3') | Reference |
| 12S | AcMDB07 | GCCTATATACCGCCGTCG | GTACACTTACCATGTTACGACTT | Bylemans et al, 2018 |
| 16S | 16S2 | GACCCTATGGAGCTTTAGAC | CGCTGTTATCCCTADRGTAACT | DiBattista et al, 2017 |
| 12S | Ac12s | ACTGGGATTAGATACCCCACTATG | GAGAGTGACGGGCGGTGT | Evans et al, 2015 |
| 16S | Ac16s | CCTTTTGCATCATGATTTAGC | CAGGTGGCTGCTTTTAGGC | Evans et al, 2015 |
| 12S | Am12s | AGCCACCGCGGTTATACG | CAAGTCCTTTGGGTTTTAAGC | Evans et al, 2015 |
| 16S | Ve16s | CGAGAAGACCCTATGGAGCTTA | AATCGTTGAACAAACGAACC | Evans et al, 2015 |
| COI | LCO-1490/HCO-2198 | TAAACTTCAGGGTGACCAAAAAATCA | TAAACTTCAGGGTGACCAAAAAATCA | Folmer et al, 1994 |
| COI | VF1d/VR1d | TTCTCAACCAACCACAARGAYATYGG | TAGACTTCTGGGTGGCCRAARAAYCA | Ivanova et al, 2006 |
| COI | VF1i/VR1i | TTCTCAACCAACCAIAAIGAIATIGG | TGITTITTIGGICACCCAGAAGTCTA | Ivanova et al, 2006 |
| COI | COI-1 | TTCTCCACCAACCACAARGAYATYGG | CACCTCAGGGTGTCCGAARAAYCARAA | Ivanova et al, 2007 |
| 12S | 12S\_F1/12S\_R1 | ACTGGGATTAGATACCCC | TAGAACAGGCTCCTCTAG | Kelly et al, 2014 |
| 16S | L2513/H2714 | GCCTGTTTACCAAAAACATCAC | CTCCATAGGGTCTTCTCGTCTT | Kitano et al, 2007 |
| CytB | CytB\_L14841/CytB\_15149R | AAAAACCACCGTTGTTATTCAACTA | GCDCCTCARAATGAYATTTGTCCTCA | Kocher et al, 1989 |
| 18S | Fish\_18S\_1F/Fish\_18S\_3R | GAATCAGGGTTCGATTCC | CAACTACGAGCTTTTTAACTGC | MacDonald et al, 2014 |
| 16S | 16S-Fish | AGCGYAATCACTTGTCTYTTAA | CRBGGTCGCCCCAACCRAA | McInnes et al, 2017 |
| CytB | L14912-CYB/H15149-CYB | TTCCTAGCCATACAYTAYAC | GGTGGCKCCTCAGAAGGACATTTGKCCYCA | Miya et al, 2000 |
| 12S | MiFish-U | GTCGGTAAAACTCGTGCCAGC | CATAGTGGGGTATCTAATCCCAGTTTG | Miya et al, 2015 |
| 16S | 16S | CGCCTGTTTATCAAAAACAT | CCGGTCTGAACTCAGATCACGT | Palumbi et al, 1996 |
| 16S | 16S1 | CGAGAAGACCCTWTGGAGCTTIAG | GGTCGCCCCAACCRAAG | Shaw et al, 2016 |
| CytB | Fish2CBL/Fish2bCBR | ACAACTTCACCCCTGCAAAC | GATGGCGTAGGCAAACAAGA | Thomsen et al, 2012 |
| CytB | Fish2degCBL/Fish2CBR | ACAACTTCACCCCTGCRAAY | GATGGCGTAGGCAAATAGGA | Thomsen et al, 2012 |
| CytB | FishCB | TCCTTTTGAGGCGCTACAGT | GGAATGCGAAGAATCGTGTT | Thomsen et al, 2012 |
| 12S | Teleo | ACACCGCCCGTCACTCT | CTTCCGGTACACTTACCATG | Valentini et al, 2016 |
| COI | FishF1/FishR1 | TCAACCAACCACAAAGACATTGGCAC | TAGACTTCTGGGTGGCCAAAGAATCA | Ward et al, 2005 |
| COI | FishF2/FishR2 | TCGACTAATCATAAAGATATCGGCAC | ACTTCAGGGTGACCGAAGAATCAGAA | Ward et al, 2005 |

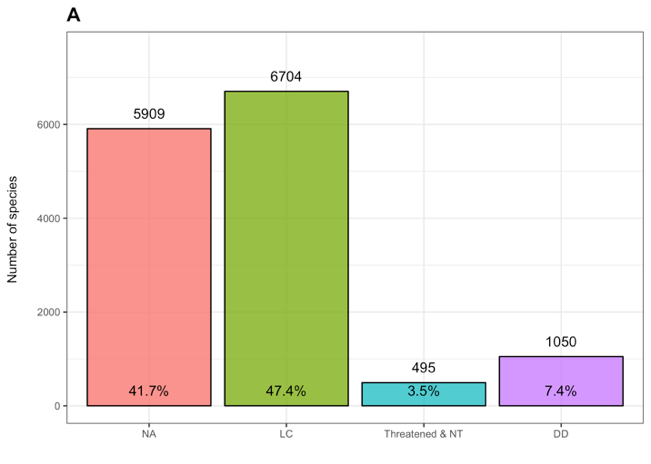
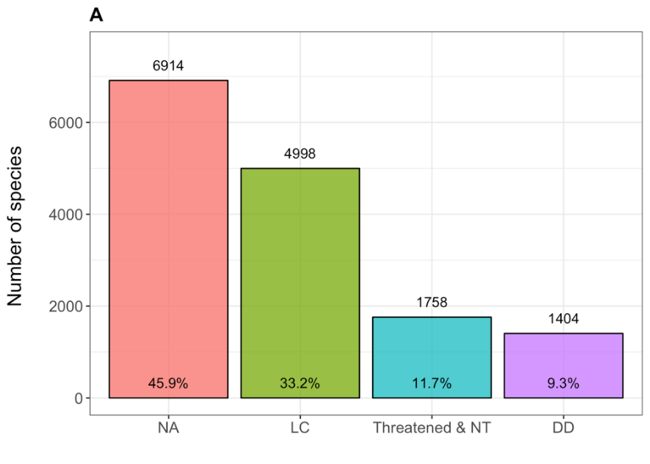
**Table S2**: All primers pairs with their percentage of coverage in freshwater, marine ecosystems and both, specifying whether they were kept for analysis or not. Percentages of sequenced species are based on the total number of freshwater (16,453), marine (15,975), or global bony fish species (32,525) from the fishbase database.

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Marker** | **Primer\_name** | **Number\_seq\_freshwater** | **Percent\_seq\_freshwater  (%)** | **Number\_seq\_marine** | **Percent\_seq\_marine  (%)** | **Number\_seq\_global** | **Percent\_seq\_global (%)** | **Kept** |
| 12S | Bylemans\_12S | 2248 | 14 | 1990 | 12 | 4253 | 13 | Yes |
| 16S | DiBattista\_16S | 3323 | 20 | 4733 | 30 | 8073 | 25 | Yes |
| 12S | Evans\_12S\_ac12s | 2072 | 13 | 2074 | 13 | 4159 | 13 | Yes |
| 16S | Evans\_16S\_ac16s | 2182 | 13 | 1542 | 10 | 3735 | 11 | Yes |
| 12S | Evans\_12S\_am12s | 2093 | 13 | 2062 | 13 | 4165 | 13 | Yes |
| 16S | Evans\_16S\_ve16s | 206 | 1 | 214 | 1 | 426 | 1 | No |
| COI | Ianova\_COI\_VF1d | 281 | 2 | 386 | 2 | 669 | 2 | No |
| COI | Ianova\_COI\_vf1d | 1473 | 9 | 1308 | 8 | 2789 | 9 | Yes |
| 12S | Kelly\_12S | 2155 | 13 | 2373 | 15 | 4541 | 14 | Yes |
| 16S | Kitano\_16S | 2163 | 13 | 2053 | 13 | 4228 | 13 | Yes |
| CYTB | Kocher\_CYTB | 1279 | 8 | 974 | 6 | 2258 | 7 | Yes |
| 18S | Macdonalds\_18S | 260 | 2 | 210 | 1 | 471 | 1 | No |
| 16S | Mcinnes\_16S | 4333 | 26 | 4721 | 30 | 9071 | 28 | Yes |
| CYTB | Miya\_CYTB | 2533 | 15 | 1806 | 11 | 4344 | 13 | Yes |
| 12S | Miya\_12S | 2068 | 13 | 1992 | 12 | 4070 | 13 | Yes |
| 16S | Palumbi\_16S | 1800 | 11 | 1488 | 9 | 3296 | 10 | Yes |
| 16S | Shaw\_16S | 4336 | 26 | 4786 | 30 | 9139 | 28 | Yes |
| CYTB | Thomsen\_CYTB\_2cbl | 2113 | 13 | 981 | 6 | 3097 | 10 | Yes |
| CYTB | Thomsen\_CYTB\_2deg | 3059 | 19 | 1629 | 10 | 4696 | 14 | Yes |
| CYTB | Thomsen\_CYTB\_cb | 2562 | 16 | 706 | 4 | 3275 | 10 | Yes |
| 12S | Valentini\_12S | 2310 | 14 | 1918 | 12 | 4242 | 13 | Yes |
| COI | Ward\_COI\_f1 | 1156 | 7 | 809 | 5 | 1970 | 6 | Yes |
| COI | Ward\_COI\_f2 | 287 | 2 | 502 | 3 | 790 | 2 | No |

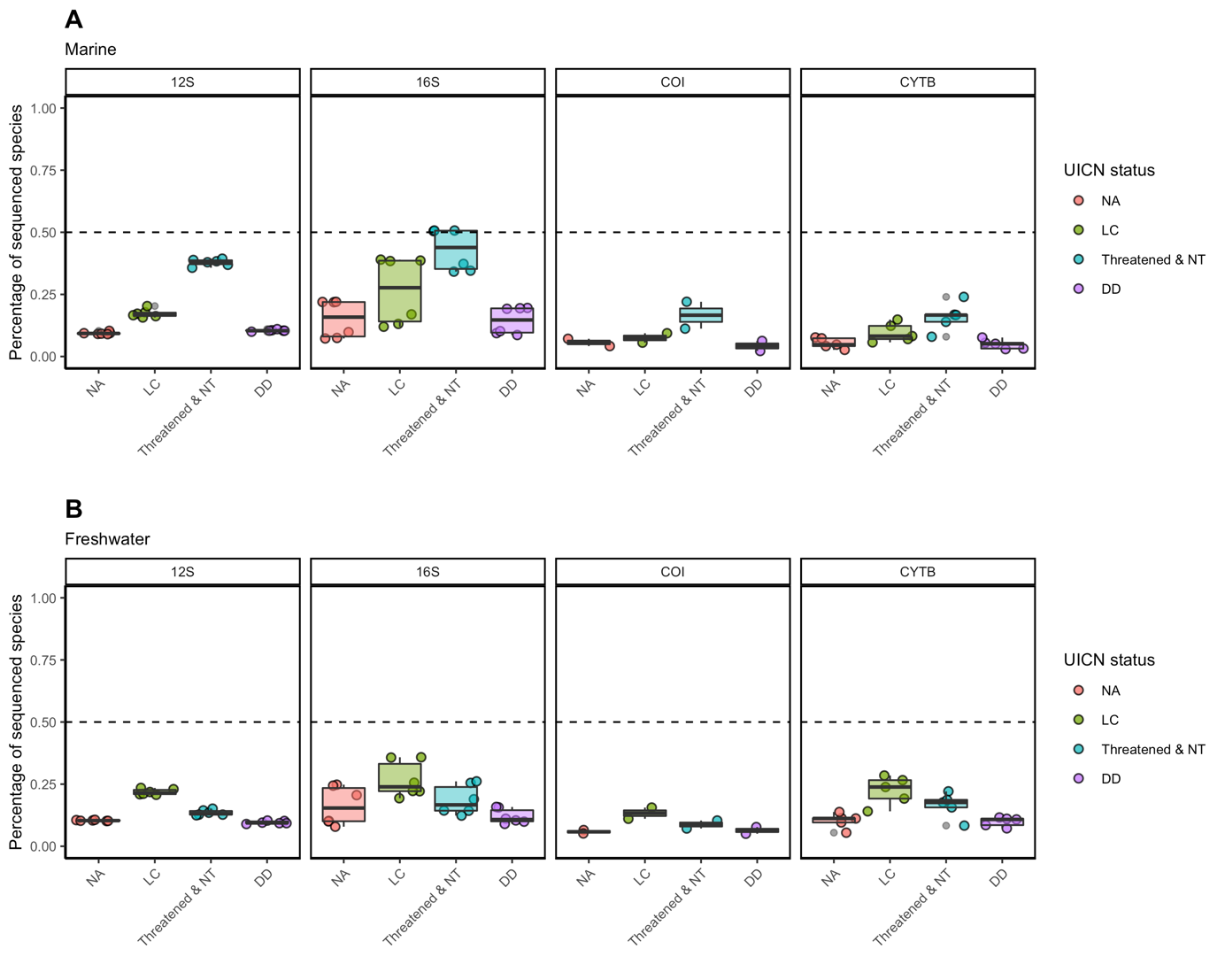


**Fig S1:** Relationship between area of polygons and the number of species for (A) freshwater with a log scale displayed for area and (B) marine ecosystems at the scale of ecoregions.

**B**



**Fig S2**: Repartition of species within IUCN categories for (A) freshwater and (B) marine species. The number above the bars and the percentage indicate the number of species in the category and the percentage of sequenced species respectively, with NA for unevaluated species (red), DD for Data Deficient (purple), LC for Least Concern (green) and Threatened & NT (blue) status among IUCN classification (http://www.iucnredlist.org.). The threatened category includes the categories Vulnerable (VU), Endangered (E) and CRitically endangered (CR).



**Fig. S3.** Percentage of coverage in each IUCN category in genetics database according to the genetic marker and IUCN status for (A) marine species and (B) freshwater species, with NA for unevaluated species (red), DD for Data Deficient (purple), LC for Least Concern (green) and Threatened & NT (blue) status among IUCN classification (http://www.iucnredlist.org.). The threatened category includes the categories Vulnerable (VU), Endangered (E) and CRitically endangered (CR). Each dot represents one primer pair.

**AD**

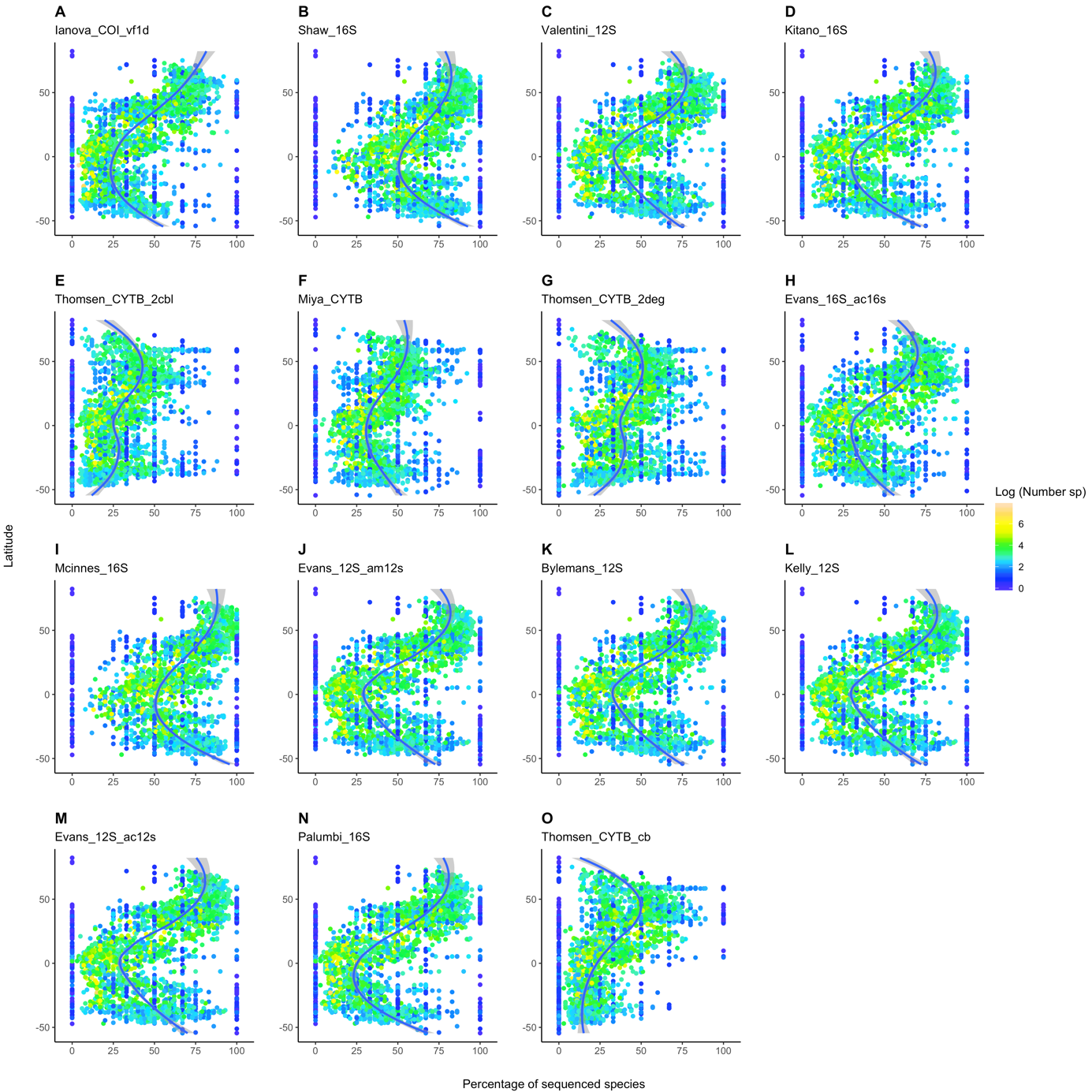
**B**

|  |  |
| --- | --- |
|  |  |
|  |  |

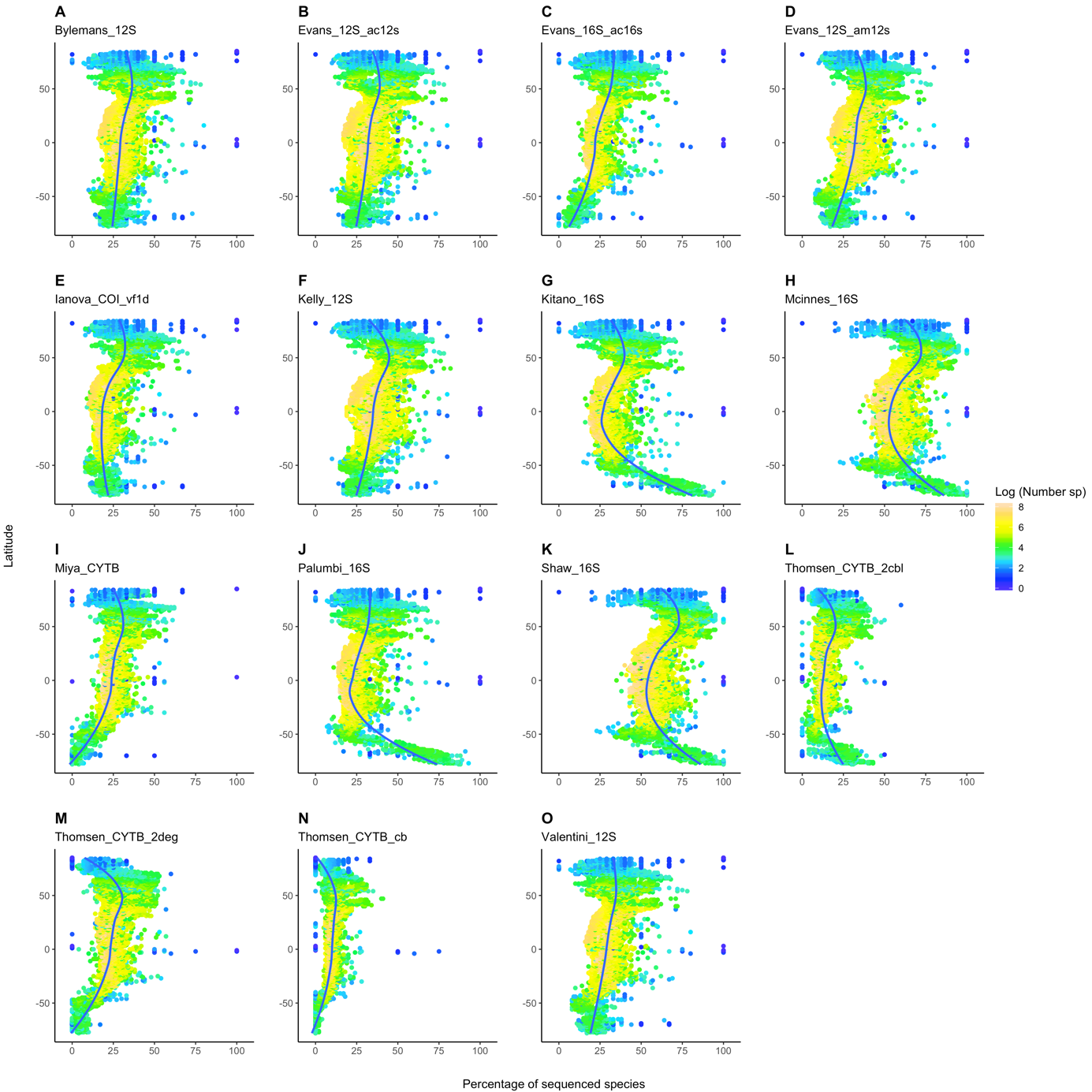
**C**

**D**

**Fig. S4**. Number of threatened species for (A) freshwater and (B) marine systems and the percentage of threatened species remaining to be sequenced for (C) freshwater and (D) marine systems.



**Fig. S5.** Percentage of species coverage in genetic database for freshwater fish species according to latitude and total species richness for the 15 markers not displayed in main (A-O).



**Fig. S6.** Percentage of species coverage in genetic database for marine fish species according to latitude and total species richness for the 15 markers not displayed in main (A-O).