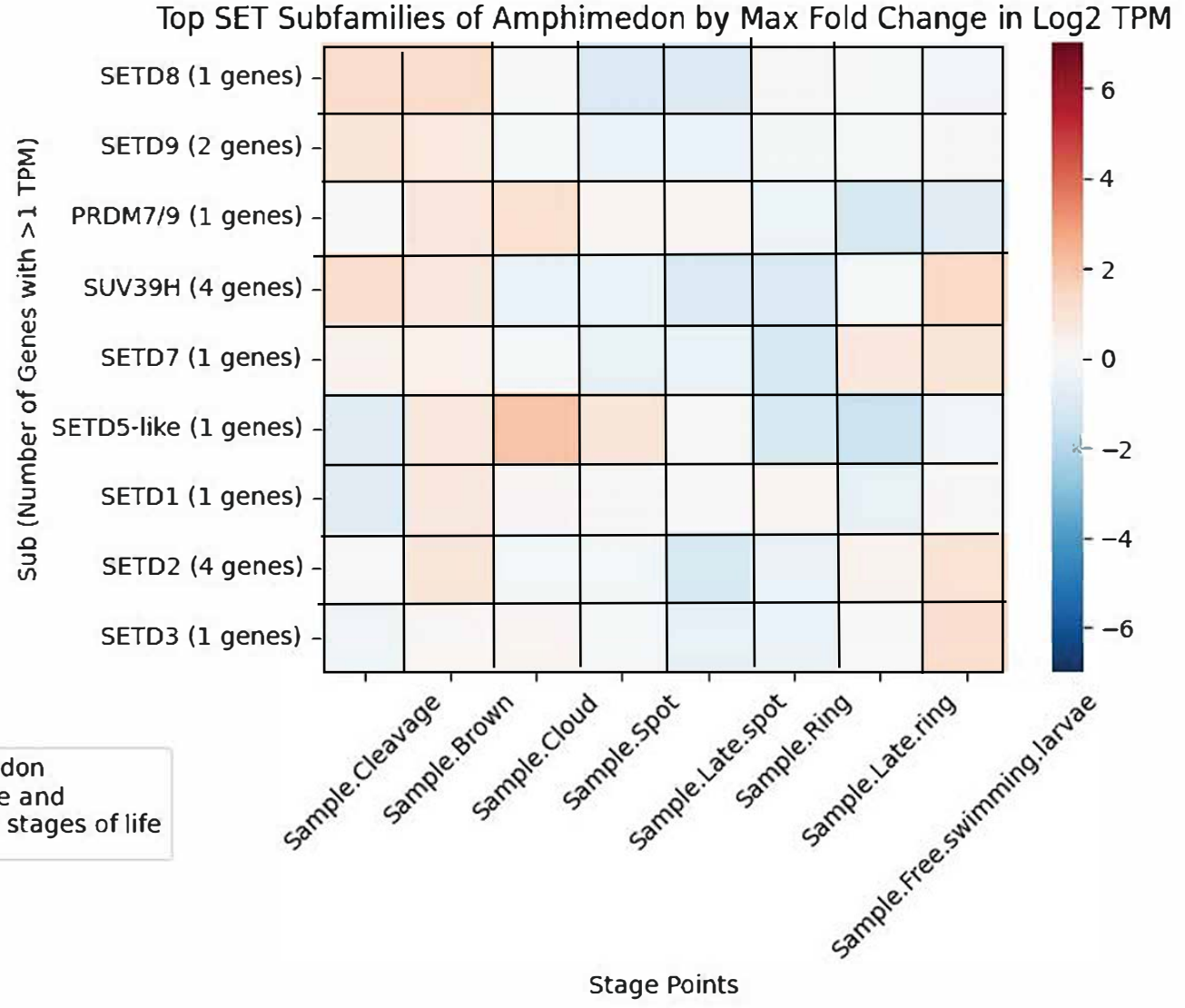


Fig 1. SET Subfamilies of Amphimedon transformed using maxfold change and ranked by prevalance within early stages of life

| Subfamily_CDDID | Gene_Name_y | Sample.Cleavage | Sample.Brown | Sample.Cloud | Sample.Spot | Sample.Late.spot | Sample.Ring | Sample.Late.ring | Sample.Free.swimming.larvae |
|---|---|---|---|---|---|---|---|---|---|
| cd10518 | SETD1 | 126.75200000000001 | 171.633 | 192.154 | 144.292 | 100.847 | 116.33699999999999 | 163.44299999999998 | 353.281 |
| cd10527 | SETD3 | 43.738 | 39.011 | 21.214 | 16.527 | 16.517 | 20.052 | 20.881999999999998 | 25.26 |
| cd10528 | SETD8 | 3.977 | 12.175 | 30.376 | 13.543 | 8.059 | 3.072 | 2.193 | 5.826 |
| cd10529 | SETD5-like | 67.503 | 69.147 | 46.762 | 35.163 | 38.088 | 22.514 | 87.295 | 92.778 |
| cd10530 | SETD7 | 5.983 | 12.001 | 5.386 | 5.458 | 2.763 | 4.355 | 8.58 | 13.334 |
| cd10531 | SETD2 | 49.643 | 147.265 | 101.035 | 92.163 | 88.467 | 102.043 | 59.422000000000004 | 94.57600000000001 |
| cd10534 | PRDM7/9 | 8.775 | 8.485 | 3.216 | 1.344 | 1.493 | 3.618 | 3.014 | 2.641 |
| cd10537 | SETD9 | 27.166 | 21.343 | 8.657 | 8.719 | 6.0089999999999995 | 5.787 | 10.788 | 32.001 |
| cd10538 | SUV39H | 22.654 | 41.183 | 48.65 | 27.751 | 27.464 | 17.308 | 10.425 | 13.015 |

| CDD_ID | Gene Name | Subfamily CDD_ID |
|---|---|---|
| cd08161 | SET | cd08161 |
| cd10518 | SETD1-like | cd10518 |
| cd19169 | SETD1 | cd10518 |
| cd19204 | SETD1A | cd10518 |
| cd19205 | SETD1B | cd10518 |
| cd19170 | KMT2A/2B | cd10518 |
| cd19206 | KMT2A | cd10518 |
| cd19207 | KMT2B | cd10518 |
| cd19171 | KMT2C/2D | cd10518 |
| cd19208 | KMT2C | cd10518 |
| cd19209 | KMT2D | cd10518 |
| cd20072 | SET1 | cd10518 |
| cd10519 | EZH | cd10519 |
| cd19168 | EZH-like | cd10519 |
| cd19217 | EZH1 | cd10519 |
| cd19218 | EZH2 | cd10519 |
| cd10522 | LegAS4-like | cd10522 |
| cd10524 | Suv4-20-like | cd10524 |
| cd19184 | KMT5B | cd10524 |
| cd19185 | KMT5C | cd10524 |
| cd19186 | Suv4-20 | cd10524 |
| cd10527 | LSMT | cd10527 |
| cd19176 | SETD3 | cd10527 |
| cd19177 | SETD4 | cd10527 |
| cd19178 | SETD6 | cd10527 |
| cd19179 | RBCMT | cd10527 |
| cd19180 | SpSET10-like | cd10527 |
| cd10528 | SETD8 | cd10528 |
| cd10529 | SETD5-like | cd10529 |
| cd19181 | SETD5 | cd10529 |
| cd19182 | KMT2E | cd10529 |
| cd19183 | SpSET3-like | cd10529 |
| cd10530 | SETD7 | cd10530 |
| cd10531 | SETD2-like | cd10531 |
| cd19172 | SETD2 | cd10531 |
| cd19173 | NSD | cd10531 |
| cd19210 | NSD1 | cd10531 |
| cd19211 | NSD2 | cd10531 |
| cd19212 | NSD3 | cd10531 |
| cd19174 | ASH1L | cd10531 |
| cd19175 | ASHR3-like | cd10531 |
| cd10534 | PRDM-like | cd10534 |
| cd10520 | PRDM17 | cd10534 |
| cd19187 | PRDM1 | cd10534 |
| cd19188 | PRDM2 | cd10534 |
| cd19189 | PRDM4 | cd10534 |
| cd19190 | PRDM5 | cd10534 |
| cd19191 | PRDM6 | cd10534 |
| cd19192 | PRDM8 | cd10534 |
| cd19193 | PRDM7/9 | cd10534 |

| | | |
|---|---|---|
| cd19194 | PRDM10 | cd10534 |
| cd19195 | PRDM11 | cd10534 |
| cd19196 | PRDM12 | cd10534 |
| cd19197 | PRDM13 | cd10534 |
| cd19198 | PRDM14 | cd10534 |
| cd19199 | PRDM15 | cd10534 |
| cd19200 | PRDM16/3 | cd10534 |
| cd19213 | PRDM16 | cd10534 |
| cd19214 | PRDM3 | cd10534 |
| cd19201 | ZFPM | cd10534 |
| cd19215 | ZFPM1 | cd10534 |
| cd19216 | ZFPM2 | cd10534 |
| cd10537 | SETD9 | cd10537 |
| cd10538 | SETDB-like | cd10538 |
| cd10541 | SETDB | cd10538 |
| cd10517 | SETDB1 | cd10538 |
| cd10523 | SETDB2 | cd10538 |
| cd10542 | SUV39H | cd10538 |
| cd10525 | SUV39H1 | cd10538 |
| cd10532 | SUV39H2 | cd10538 |
| cd19473 | DIM5-like | cd10538 |
| cd20073 | Clr4-like | cd10538 |
| cd10543 | EHMT | cd10538 |
| cd10533 | EHMT2 | cd10538 |
| cd10535 | EHMT1 | cd10538 |
| cd10544 | SETMAR | cd10538 |
| cd10545 | AtSUVH-like | cd10538 |
| cd10545 | ATXR5/6-like | cd10545 |
| cd10540 | SpSet7-like | cd10540 |
| cd20071 | SMYD | cd20071 |
| cd10521 | SMYD5 | cd20071 |
| cd10536 | SMYD4 | cd20071 |
| cd19167 | SMYD1/2/3-like | cd20071 |
| cd10526 | SMYD1 | cd20071 |
| cd19202 | SMYD2 | cd20071 |
| cd19203 | SMYD3 | cd20071 |

For this project, I am focusing on the SET Subfamilies and their expression throughout both of the Taxa. I did this by hand since the SET subfamilies had very few actual subfamilies to annotate. I looked up the information on the NCBI website using the superfamily number that was found in the first part of the project. I then created a Google Sheets file that had each Gene's Subfamily and its CDD_ID number. I then transferred this file to a .txt file and merged it with the rest of my tables, using the same code as project one. The difference with this one was that I was now looking at the subfamilies instead of the superfamilies, filtering out ones that did not align with the ones I was observing. However, I used the Top 20 Max Fold Change so i could see how many of these genes are used within each subfamily. I think my next move is to see what each of these subfamilies do and compare that to their trends in Amphemideon.