# Supplementary Information for the independent and compreshensive evaluation of CAZyme classifiers

Emma E. M. Hobbs, Tracey M. Gloster, Leighton Pritchard

June 2023

This document contains all supplementary information for the paper: Independent and comprehensive evaluation of CAZyme classifiers (Hobbs *et al.*, 2023). The tables and figures are presented in the same order as they are referenced in the main manuscript.

## Contents

# 1 Test set composition

## SI Table 1: Genomes used to compile test sets (overleaf)

SI Table 1 lists the taxonomic classification of each genomic assembly used to generate test sets to evaluate the CAZyme classifiers. The number of CAZymes listed per species (* encompassing all strains) and per genus is taken from the July 2023 CAZy database release. In the July 2023 CAZy database release: *Streptomyces antimycoticus* (GCA_009936315.1) is classified *Brettanomyces nanus*; *Brettanomyces nanus* (GCA_011074865.2) is classified as *Brettanomyces bruxellensis UCD 2041*; and *Nibricoccus aquaticus* (GCA_002310495.1) is classified as *Verrucomicrobiota bacterium HZ-65* .

| NCBI:txid | Assembly name | Version accession | Organism | Kingdom | Group | Number of CAZymes for species in CAZy* | Number of CAZymes for genus in CAZy* |
|---|---|---|---|---|---|---|---|
| 515635 | ASM2164v1 | GCA_000021645.1 | Dictyoglomus turgidum DSM 6724 | Bacteria | Gram negative | 102 | 208 |
| 509190 | ASM9228v1 | GCA_000092285.1 | Caulobacter segnis ATCC 21756 | Bacteria | Gram negative; a-proteobacteria | 586 | 1975 |
| 192 | ASM131501v1 | GCA_001315015.1 | Azospirillum brasilense | Bacteria | Gram negative; a-proteobacteria | 1550 | 3057 |
| 661488 | ASM797018v1 | GCA_007970185.1 | Pseudobacter ginsenosidimutans | Bacteria | Gram negative; CFB group bacteria | 237 | 237 |
| 562 | ASM522158v1 | GCA_005221585.1 | Escherichia coli | Bacteria | Gram negative; E. coli | 110387 | 122982 |
| 562 | ASM522190v1 | GCA_005221905.1 | Escherichia coli | Bacteria | Gram negative; E. coli | 110387 | 122982 |
| 561229 | ASM2356v1 | GCA_000023565.1 | Dickeya chrysanthemi Ech1591 | Bacteria | Gram negative; Enterobacteriaceae | 226 | 6498 |
| 1334564 | ASM82877v1 | GCA_000828775.1 | Serratia marcescens SM39 | Bacteria | Gram negative; Enterobacteriaceae | 819 | 17744 |
| 571 | ASM290639v1 | GCA_002906395.1 | Klebsiella oxytoca | Bacteria | Gram negative; Enterobacteriaceae | 5943 | 118339 |
| 1940567 | ASM340313v1 | GCA_003403135.1 | Dickeya dianthicola | Bacteria | Gram negative; Enterobacteriaceae | 1811 | 6498 |
| 61645 | ASM394076v1 | GCA_003940765.1 | Enterobacter asburiae | Bacteria | Gram negative; Enterobacteriaceae | 3772 | 38478 |
| 1134687 | ASM1009300v1 | GCA_010093005.1 | Klebsiella michiganensis | Bacteria | Gram negative; Enterobacteriaceae | 9145 | 118339 |
| 548 | ASM1160472v1 | GCA_011604725.1 | Klebsiella aerogenes | Bacteria | Gram negative; Enterobacteriaceae | 6884 | 118339 |
| 615 | ASM1342615v1 | GCA_013426155.1 | Serratia marcescens | Bacteria | Gram negative; Enterobacteriaceae | 8192 | 17744 |
| 59203 | 31885_G02 | GCA_900635675.1 | Salmonella enterica subsp. arizonae | Bacteria | Gram negative; Enterobacteriaceae | 48446 | 51679 |
| 498211 | ASM1922v1 | GCA_000019225.1 | Cellvibrio japonicus Ueda107 | Bacteria | Gram negative; g-proteobacteria | 912 | 2091 |
| 1137651 | ASM34922v1 | GCA_000349225.1 | Xanthomonas citri subsp. citri Aw12879 | Bacteria | Gram negative; g-proteobacteria | 18260 | 70327 |
| 1308541 | ASM81688v1 | GCA_000816885.1 | Xanthomonas citri subsp. citri A306 | Bacteria | Gram negative; g-proteobacteria | 18260 | 70327 |
| 1583341 | PCPL58T | GCA_900074915.1 | Pseudomonas cerasi | Bacteria | Gram negative; g-proteobacteria | 304 | 97188 |
| 718 | 55685_B01 | GCA_900638075.1 | Actinobacillus equuli | Bacteria | Gram negative; g-proteobacteria | 1688 | 4578 |
| 1637999 | ASM97276v1 | GCA_000972765.1 | Verrucomicrobia bacterium IMCC26134 | Bacteria | Gram negative; verrucomicrobia | 375 | 375 |
| 2026799 | ASM231049v1 | GCA_002310495.1 | Nibricoccus aquaticus | Bacteria | Gram negative; verrucomicrobia | 284 | 284 |
| 203119 | ASM1586v1 | GCA_000015865.1 | Acetivibrio thermocellus ATCC 27405 | Bacteria | Gram positive; firmicutes | 828 | 1111 |
| 394503 | ASM2206v1 | GCA_000022065.1 | Ruminiclostridium cellulolyticum H10 | Bacteria | Gram positive; firmicutes | 168 | 524 |
| 720554 | ASM23708v1 | GCA_000237085.1 | Acetivibrio clariflavus DSM 19732 | Bacteria | Gram positive; firmicutes | 145 | 1111 |
| 1520 | ASM83310v2 | GCA_000833105.2 | Clostridium beijerinckii | Bacteria | Gram positive; firmicutes | 1427 | 28279 |
| 1292358 | ASM83514v1 | GCA_000835145.1 | Bacillus amyloliquefaciens KHG19 | Bacteria | Gram positive; firmicutes | 7463 | 111030 |
| 36745 | ASM200330v1 | GCA_002003305.1 | Clostridium saccharoperbutylacetonicum | Bacteria | Gram positive; firmicutes | 453 | 28279 |
| 1352 | ASM202504v1 | GCA_002025045.1 | Enterococcus faecium | Bacteria | Gram positive; firmicutes | 11568 | 27760 |
| 304207 | ASM869410v1 | GCA_008694105.1 | Schleiferilactobacillus harbinensis | Bacteria | Gram positive; firmicutes | 435 | 435 |
| 37734 | ASM970734v1 | GCA_009707345.1 | Enterococcus casseliflavus | Bacteria | Gram positive; firmicutes | 1368 | 22760 |
| 1429244 | ASM50720v2 | GCA_000507205.2 | Paenibacillus polymyxa CR1 | Bacteria | Gram positive; firmicutes | 8515 | 36966 |
| 1039 | ASM1479206v1 | GCA_014792065.1 | Bacillus amyloliquefaciens | Bacteria | Gram positive; firmicutes | 7463 | 111030 |
| 2665646 | ASM1640612v1 | GCA_016406125.1 | Alicyclobacillus sp. SO9 | Bacteria | Gram positive; firmicutes | 442 | 1232 |
| 479432 | ASM2486v1 | GCA_000024865.1 | Streptosporangium roseum DSM 43021 | Bacteria | Gram positive; high G+C | 257 | 605 |
| 749414 | ASM9238v1 | GCA_000092385.1 | Streptomyces bingchenggensis BCW-1 | Bacteria | Gram positive; high G+C | 387 | 100828 |
| 212767 | ASM32856v1 | GCA_000328565.1 | Mycobacterium sp. JS623 | Bacteria | Gram positive; high G+C | 9195 | 30430 |
| 284038 | ASM993631v1 | GCA_009936315.1 | Streptomyces antimycoticus | Bacteria | Gram positive; high G+C | 140 | 100828 |
| 228602 | ASM1180114v1 | GCA_011801145.1 | Nocardia arthritidis | Bacteria | Gram positive; high G+C | 180 | 3932 |
| 2704468 | ASM1408410v1 | GCA_014084105.1 | Streptacidiphilus sp. P02-A3a | Bacteria | Gram positive; high G+C | 449 | 499 |
| 332648 | ASM14353v4 | GCA_000143535.4 | Botrytis cinerea B05.10 | Eukaryote | Ascomycete fungi | 1309 | 1371 |
| 403677 | ASM159280v2 | GCA_001592805.2 | Peltaster fructicola | Eukaryote | Ascomycete fungi | 266 | 266 |
| 318829 | ASM434696v1 | GCA_004346965.1 | Pyricularia oryzae | Eukaryote | Ascomycete fungi | 1757 | 1825 |
| 73501 | ASM808049v1 | GCA_008080495.1 | Cordyceps militaris | Eukaryote | Ascomycete fungi | 323 | 356 |
| 227321 | ASM901741v1 | GCA_009017415.1 | Aspergillus flavus | Eukaryote | Ascomycete fungi | 2516 | 7327 |
| 660027 | ASM1308505v1 | GCA_013085055.1 | Fusarium oxysporum Fo47 | Eukaryote | Ascomycete fungi | 1827 | 8833 |
| 500148 | ASM1342620v1 | GCA_013426205.1 | Metarhizium brunneum | Eukaryote | Ascomycete fungi | 393 | 490 |
| 36651 | ASM1676781v1 | GCA_016767815.1 | Penicillium digitatum | Eukaryote | Ascomycete fungi | 345 | 1037 |
| 182096 | AchevalieriM1_assembly01 | GCA_016861735.1 | Aspergillus chevalieri | Eukaryote | Ascomycete fungi | 311 | 7327 |
| 101028 | ASM1695230v1 | GCA_016952305.1 | Fusarium pseudograminearum | Eukaryote | Ascomycete fungi | 1019 | 8833 |
| 5516 | ASM1695235v1 | GCA_016952355.1 | Fusarium culmorum | Eukaryote | Ascomycete fungi | 485 | 8833 |
| 2747967 | ASM2310122v1 | GCA_023101225.1 | Fusarium solani-melongenae CRI 24-3 | Eukaryote | Ascomycete fungi | 681 | 8833 |
| 5059 | ASM1478422v2 | GCA_014784225.2 | Aspergillus flavus CA14 | Eukaryote | Ascomycete fungi | 2516 | 7327 |
| 5499 | Cfulv_R5_v5 | GCA_020509005.2 | Fulvia fulva Race5_Kim | Eukaryote | Ascomycete fungi | 576 | 576 |
| 63577 | ASM2064779v1 | GCA_020647795.1 | Trichoderma atroviride P1 | Eukaryote | Ascomycete fungi | 586 | 3361 |
| 101201 | ASM2064786v1 | GCA_020647865.1 | Trichoderma asperellum FT101 | Eukaryote | Ascomycete fungi | 550 | 3361 |
| 1491479 | ASM1956561v1 | GCA_019565615.1 | Trichoderma simmonsii GH-Sj1 | Eukaryote | Ascomycete fungi | 483 | 3361 |
| 170446 | ASM1690657v1 | GCA_016906575.1 | Ceratobasidium sp. AG-Ba | Eukaryote | Basidiomycota fungi | 1287 | 1918 |
| 284590 | ASM251v1 | GCA_000002515.1 | Kluyveromyces lactis | Eukaryote | Budding yeasts | 296 | 730 |
| 573826 | ASM2694v1 | GCA_000026945.1 | Candida dubliniensis CD36 | Eukaryote | Budding yeasts | 146 | 3700 |
| 284811 | ASM9102v4 | GCA_000091025.4 | Eremothecium gossypii ATCC 10895 | Eukaryote | Budding yeasts | 270 | 477 |
| 796027 | ASM164002v2 | GCA_001640025.2 | Sugiyamaella lignohabitans | Eukaryote | Budding yeasts | 150 | 150 |
| 4911 | ASM185444v2 | GCA_001854445.2 | Kluyveromyces marxianus | Eukaryote | Budding yeasts | 434 | 730 |
| 1365886 | ASM198439v2 | GCA_001984395.2 | Zygosaccharomyces parabailii | Eukaryote | Budding yeasts | 220 | 450 |
| 498019 | ASM301371v2 | GCA_003013715.2 | [Candida] auris | Eukaryote | Budding yeasts | 657 | 3700 |
| 4909 | ASM305444v1 | GCA_003054445.1 | Pichia kudriavzevii | Eukaryote | Budding yeasts | 188 | 229 |
| 2163413 | ASM421770v1 | GCA_004217705.1 | Metschnikowia aff. pulcherrima | Eukaryote | Budding yeasts | 135 | 144 |
| 28985 | ASM799369v1 | GCA_007993695.1 | Kluyveromyces lactis | Eukaryote | Budding yeasts | 296 | 730 |
| 498019 | ASM827514v1 | GCA_008275145.1 | [Candida] auris | Eukaryote | Budding yeasts | 657 | 3700 |
| 36911 | ASM949811v1 | GCA_009498115.1 | Clavispora lusitaniae | Eukaryote | Budding yeasts | 686 | 686 |
| 13502 | ASM1107486v2 | GCA_011074865.2 | Brettanomyces nanus | Eukaryote | Budding yeasts | 138 | 279 |
| 5007 | ASM1107488v2 | GCA_011074885.2 | Brettanomyces bruxellensis | Eukaryote | Budding yeasts | 138 | 279 |
| 5478 | ASM1421772v1 | GCA_014217725.1 | Nakaseomyces glabratus | Eukaryote | Budding yeasts | 2152 | 2252 |
| 4652 | ASM1449061v1 | GCA_014490615.1 | Yarrowia lipolytic | Eukaryote | Budding yeasts | 407 | 407 |
| 230603 | ASM2755758v1 | GCA_027557585.1 | Saccharomyces uvarum CBS7001 | Eukaryote | Budding yeasts | 541 | 17183 |
| 296587 | ASM9098v2 | GCA_000090985.2 | Micromonas commoda | Eukaryote | Green algae | 150 | 152 |
| 436017 | ASM9206v1 | GCA_000092065.1 | Ostreococcus lucimarinus CCE9901 | Eukaryote | Green algae | 116 | 246 |
| 1764295 | ASM785969v1 | GCA_007859695.1 | Chloropicon primus | Eukaryote | Green algae | 443 | 443 |
| 6239 | WBcel235 | GCA_000002985.3 | Caenorhabditis elegans BRISTOL N2 | Eukaryote | Nematodes | 1340 | 1880 |
| 573729 | ASM22609v1 | GCA_000226095.1 | Thermothelomyces thermophilus ATCC 42464 | Eukaryote | Thermophile | 400 | 401 |

## SI Table 2: Coverage of CAZy families (overleaf)

SI Table 2 lists the number of unique NCBI protein version accessions associated with each CAZy family in the CAZy database July 2023 release, as well as the total number and percentage of proteins from each CAZy family included across the test sets. CAZy families are grouped by their respective parent CAZy class.

## GH

| Family | Sample Size | Population | Percentage in the test sets |
|---|---|---|---|
| GH0 | 119 | 33642 | 0.35 |
| GH1 | 138 | 61659 | 0.22 |
| GH2 | 123 | 40275 | 0.31 |
| GH3 | 213 | 58104 | 0.37 |
| GH4 | 66 | 26324 | 0.25 |
| GH5 | 179 | 28654 | 0.62 |
| GH6 | 14 | 3896 | 0.36 |
| GH7 | 16 | 10823 | 0.15 |
| GH8 | 21 | 9649 | 0.22 |
| GH9 | 36 | 6420 | 0.56 |
| GH10 | 40 | 8285 | 0.48 |
| GH11 | 22 | 2753 | 0.8 |
| GH12 | 20 | 2608 | 0.77 |
| GH13 | 323 | 162804 | 0.2 |
| GH14 | 4 | 1335 | 0.3 |
| GH15 | 46 | 13136 | 0.35 |
| GH16 | 148 | 18481 | 0.8 |
| GH17 | 82 | 8018 | 1.02 |
| GH18 | 203 | 38371 | 0.53 |
| GH19 | 17 | 15540 | 0.11 |
| GH20 | 28 | 18108 | 0.15 |
| GH21 | 0 | 0 | 0 |
| GH22 | 2 | 1842 | 0.11 |
| GH23 | 164 | 120343 | 0.14 |
| GH24 | 51 | 24476 | 0.21 |
| GH25 | 19 | 18893 | 0.1 |
| GH26 | 18 | 5238 | 0.34 |
| GH27 | 29 | 4715 | 0.62 |
| GH28 | 68 | 16335 | 0.42 |
| GH29 | 21 | 11212 | 0.19 |
| GH30 | 27 | 6225 | 0.43 |
| GH31 | 101 | 23886 | 0.42 |
| GH32 | 65 | 22942 | 0.28 |
| GH33 | 17 | 10310 | 0.16 |
| GH34 | 0 | 136977 | 0 |
| GH35 | 37 | 9012 | 0.41 |
| GH36 | 40 | 13045 | 0.31 |
| GH37 | 36 | 11414 | 0.32 |
| GH38 | 38 | 10833 | 0.35 |
| GH39 | 21 | 5341 | 0.39 |
| GH40 | 0 | 0 | 0 |
| GH41 | 0 | 0 | 0 |
| GH42 | 20 | 8914 | 0.22 |
| GH43 | 160 | 35019 | 0.46 |
| GH44 | 3 | 331 | 0.91 |
| GH45 | 4 | 671 | 0.6 |
| GH46 | 6 | 1383 | 0.43 |
| GH47 | 74 | 2872 | 2.58 |
| GH48 | 4 | 1707 | 0.23 |
| GH49 | 1 | 75 | 1.33 |
| GH50 | 2 | 1640 | 0.12 |
| GH51 | 40 | 8244 | 0.49 |
| GH52 | 3 | 215 | 1.4 |
| GH53 | 18 | 5003 | 0.36 |
| GH54 | 5 | 561 | 0.89 |
| GH55 | 28 | 1827 | 1.53 |
| GH56 | 1 | 484 | 0.21 |
| GH57 | 7 | 4963 | 0.14 |
| GH58 | 0 | 223 | 0 |
| GH59 | 6 | 445 | 1.35 |
| GH60 | 0 | 0 | 0 |
| GH61 | 0 | 0 | 0 |
| GH62 | 6 | 903 | 0.66 |
| GH63 | 22 | 6166 | 0.36 |
| GH64 | 8 | 1306 | 0.61 |
| GH65 | 39 | 10961 | 0.36 |
| GH66 | 0 | 474 | 0 |
| GH67 | 13 | 1795 | 0.72 |
| GH68 | 6 | 2260 | 0.27 |
| GH69 | 0 | 0 | 0 |
| GH70 | 0 | 998 | 0 |
| GH71 | 20 | 566 | 3.53 |
| GH72 | 93 | 1537 | 6.05 |
| GH73 | 38 | 37364 | 0.1 |
| GH74 | 10 | 1371 | 0.73 |
| GH75 | 11 | 601 | 1.83 |
| GH76 | 72 | 4365 | 1.65 |
| GH77 | 28 | 16961 | 0.17 |
| GH78 | 42 | 7733 | 0.54 |
| GH79 | 9 | 1180 | 0.76 |
| GH80 | 0 | 17 | 0 |
| GH81 | 36 | 1557 | 2.31 |
| GH82 | 0 | 119 | 0 |
| GH83 | 0 | 8118 | 0 |
| GH84 | 5 | 3049 | 0.16 |
| GH85 | 4 | 1699 | 0.24 |
| GH86 | 0 | 256 | 0 |
| GH87 | 5 | 1612 | 0.31 |
| GH88 | 13 | 5354 | 0.24 |
| GH89 | 5 | 1925 | 0.26 |
| GH90 | 0 | 545 | 0 |
| GH91 | 2 | 607 | 0.33 |
| GH92 | 42 | 12989 | 0.32 |
| GH93 | 9 | 582 | 1.55 |
| GH94 | 14 | 5566 | 0.25 |
| GH95 | 27 | 5533 | 0.49 |
| GH96 | 0 | 15 | 0 |
| GH97 | 8 | 5163 | 0.15 |
| GH98 | 2 | 537 | 0.37 |
| GH99 | 1 | 463 | 0.22 |
| GH100 | 0 | 884 | 0 |
| GH101 | 1 | 1053 | 0.09 |
| GH102 | 14 | 12138 | 0.12 |
| GH103 | 23 | 23647 | 0.1 |
| GH104 | 8 | 3829 | 0.21 |
| GH105 | 35 | 9288 | 0.38 |
| GH106 | 10 | 2025 | 0.49 |
| GH107 | 0 | 36 | 0 |
| GH108 | 11 | 6891 | 0.16 |
| GH109 | 0 | 2192 | 0 |
| GH110 | 1 | 987 | 0.1 |
| GH111 | 0 | 30 | 0 |
| GH112 | 1 | 803 | 0.12 |
| GH113 | 5 | 1824 | 0.27 |
| GH114 | 6 | 1576 | 0.38 |
| GH115 | 16 | 1950 | 0.82 |
| GH116 | 1 | 1152 | 0.09 |
| GH117 | 0 | 675 | 0 |
| GH118 | 0 | 11 | 0 |
| GH119 | 0 | 61 | 0 |
| GH120 | 3 | 492 | 0.61 |
| GH121 | 1 | 595 | 0.17 |
| GH122 | 0 | 90 | 0 |
| GH123 | 2 | 934 | 0.21 |
| GH124 | 0 | 7 | 0 |
| GH125 | 25 | 3967 | 0.63 |
| GH126 | 4 | 1645 | 0.24 |
| GH127 | 8 | 4699 | 0.17 |
| GH128 | 16 | 973 | 1.64 |
| GH129 | 0 | 180 | 0 |
| GH130 | 17 | 4632 | 0.37 |
| GH131 | 5 | 153 | 3.27 |
| GH132 | 44 | 984 | 4.47 |
| GH133 | 21 | 1841 | 1.14 |
| GH134 | 1 | 168 | 0.6 |
| GH135 | 12 | 886 | 1.35 |
| GH136 | 2 | 1034 | 0.19 |
| GH137 | 0 | 268 | 0 |
| GH138 | 1 | 245 | 0.41 |
| GH139 | 1 | 274 | 0.36 |
| GH140 | 3 | 1217 | 0.25 |
| GH141 | 7 | 1159 | 0.6 |
| GH142 | 3 | 453 | 0.66 |
| GH143 | 1 | 231 | 0.43 |
| GH144 | 2 | 2764 | 0.07 |
| GH145 | 0 | 0 | 0 |
| GH146 | 9 | 2497 | 0.36 |
| GH147 | 0 | 224 | 0 |
| GH148 | 0 | 478 | 0 |
| GH149 | 0 | 207 | 0 |
| GH150 | 0 | 56 | 0 |
| GH151 | 0 | 236 | 0 |
| GH152 | 7 | 1459 | 0.48 |
| GH153 | 8 | 4869 | 0.16 |
| GH154 | 17 | 3926 | 0.43 |
| GH155 | 0 | 0 | 0 |
| GH156 | 0 | 63 | 0 |
| GH157 | 0 | 13 | 0 |
| GH158 | 2 | 836 | 0.24 |
| GH159 | 1 | 323 | 0.31 |
| GH160 | 0 | 23 | 0 |
| GH161 | 1 | 215 | 0.47 |
| GH162 | 1 | 34 | 2.94 |
| GH163 | 2 | 379 | 0.53 |
| GH164 | 0 | 35 | 0 |
| GH165 | 0 | 404 | 0 |
| GH166 | 1 | 1302 | 0.08 |
| GH167 | 0 | 105 | 0 |
| GH168 | 0 | 53 | 0 |
| GH169 | 0 | 9 | 0 |
| GH170 | 6 | 6798 | 0.09 |
| GH171 | 4 | 4685 | 0.09 |
| GH172 | 2 | 1715 | 0.12 |
| GH173 | 0 | 172 | 0 |
| GH174 | 0 | 94 | 0 |
| GH175 | 1 | 55 | 1.82 |
| GH176 | 7 | 3244 | 0.22 |
| GH177 | 0 | 832 | 0 |
| GH178 | 1 | 1016 | 0.1 |
| GH179 | 0 | 46 | 0 |
| GH180 | 0 | 11 | 0 |
| GH181 | 0 | 191 | 0 |
| GH182 | 0 | 263 | 0 |
| GH183 | 3 | 2574 | 0.12 |

## GT

| Family | Sample Size | Population | Percentage in the test sets |
|---|---|---|---|
| GT0 | 87 | 41681 | 0.21 |
| GT1 | 179 | 44225 | 0.4 |
| GT2 | 582 | 344651 | 0.17 |
| GT3 | 17 | 1623 | 1.05 |
| GT4 | 371 | 270182 | 0.14 |
| GT5 | 52 | 28657 | 0.18 |
| GT6 | 0 | 1015 | 0 |
| GT7 | 5 | 1583 | 0.32 |
| GT8 | 72 | 22489 | 0.32 |
| GT9 | 61 | 43940 | 0.14 |
| GT10 | 6 | 2418 | 0.25 |
| GT11 | 9 | 2498 | 0.36 |
| GT12 | 0 | 163 | 0 |
| GT13 | 3 | 493 | 0.61 |
| GT14 | 4 | 3577 | 0.11 |
| GT15 | 100 | 2021 | 4.95 |
| GT16 | 0 | 257 | 0 |
| GT17 | 3 | 778 | 0.39 |
| GT18 | 0 | 126 | 0 |
| GT19 | 22 | 17237 | 0.13 |
| GT20 | 72 | 15232 | 0.47 |
| GT21 | 14 | 2186 | 0.64 |
| GT22 | 80 | 2094 | 3.82 |
| GT23 | 1 | 1244 | 0.08 |
| GT24 | 23 | 552 | 4.17 |
| GT25 | 11 | 9647 | 0.11 |
| GT26 | 27 | 16892 | 0.16 |
| GT27 | 4 | 1661 | 0.24 |
| GT28 | 58 | 35432 | 0.16 |
| GT29 | 32 | 1967 | 1.63 |
| GT30 | 17 | 17253 | 0.1 |
| GT31 | 39 | 4990 | 0.78 |
| GT32 | 77 | 7636 | 1.01 |
| GT33 | 18 | 523 | 3.44 |
| GT34 | 39 | 1282 | 3.04 |
| GT35 | 74 | 26370 | 0.28 |
| GT36 | 0 | 0 | 0 |
| GT37 | 0 | 367 | 0 |
| GT38 | 0 | 172 | 0 |
| GT39 | 95 | 5426 | 1.75 |
| GT40 | 0 | 66 | 0 |
| GT41 | 24 | 6851 | 0.35 |
| GT42 | 0 | 498 | 0 |
| GT43 | 2 | 709 | 0.28 |
| GT44 | 6 | 1404 | 0.43 |
| GT45 | 0 | 238 | 0 |
| GT46 | 0 | 0 | 0 |
| GT47 | 7 | 2305 | 0.3 |
| GT48 | 45 | 4867 | 0.92 |
| GT49 | 1 | 508 | 0.2 |
| GT50 | 17 | 532 | 3.2 |
| GT51 | 97 | 88438 | 0.11 |
| GT52 | 2 | 1081 | 0.19 |
| GT53 | 4 | 2927 | 0.14 |
| GT54 | 0 | 342 | 0 |
| GT55 | 1 | 168 | 0.6 |
| GT56 | 13 | 5182 | 0.25 |
| GT57 | 41 | 1005 | 4.08 |
| GT58 | 24 | 741 | 3.24 |
| GT59 | 15 | 462 | 3.25 |
| GT60 | 3 | 539 | 0.56 |
| GT61 | 2 | 5176 | 0.04 |
| GT62 | 60 | 940 | 6.38 |
| GT63 | 0 | 68 | 0 |
| GT64 | 3 | 693 | 0.43 |
| GT65 | 0 | 222 | 0 |
| GT66 | 18 | 2411 | 0.75 |
| GT67 | 0 | 207 | 0 |
| GT68 | 0 | 181 | 0 |
| GT69 | 13 | 4011 | 0.32 |
| GT70 | 1 | 681 | 0.15 |
| GT71 | 100 | 1574 | 6.35 |
| GT72 | 0 | 322 | 0 |
| GT73 | 4 | 2956 | 0.14 |
| GT74 | 0 | 27 | 0 |
| GT75 | 0 | 459 | 0 |
| GT76 | 15 | 645 | 2.33 |
| GT77 | 15 | 723 | 2.07 |
| GT78 | 0 | 9 | 0 |
| GT79 | 0 | 24 | 0 |
| GT80 | 3 | 249 | 1.2 |
| GT81 | 5 | 3610 | 0.14 |
| GT82 | 0 | 414 | 0 |
| GT83 | 11 | 13629 | 0.08 |
| GT84 | 5 | 2619 | 0.19 |
| GT85 | 2 | 1204 | 0.17 |
| GT86 | 0 | 0 | 0 |
| GT87 | 6 | 4830 | 0.12 |
| GT88 | 0 | 245 | 0 |
| GT89 | 2 | 1438 | 0.14 |
| GT90 | 18 | 1462 | 1.23 |
| GT91 | 28 | 238 | 11.76 |
| GT92 | 5 | 461 | 1.08 |
| GT93 | 0 | 18 | 0 |
| GT94 | 1 | 805 | 0.12 |
| GT95 | 6 | 150 | 4 |
| GT96 | 2 | 119 | 1.68 |
| GT97 | 0 | 131 | 0 |
| GT98 | 0 | 207 | 0 |
| GT99 | 0 | 158 | 0 |
| GT100 | 0 | 442 | 0 |
| GT101 | 0 | 778 | 0 |
| GT102 | 0 | 455 | 0 |
| GT103 | 0 | 144 | 0 |
| GT104 | 2 | 3466 | 0.06 |
| GT105 | 0 | 708 | 0 |
| GT106 | 0 | 1143 | 0 |
| GT107 | 0 | 3178 | 0 |
| GT108 | 0 | 588 | 0 |
| GT109 | 0 | 73 | 0 |
| GT110 | 0 | 301 | 0 |
| GT111 | 3 | 2374 | 0.13 |
| GT112 | 0 | 577 | 0 |
| GT113 | 1 | 1531 | 0.07 |
| GT114 | 0 | 124 | 0 |
| GT115 | 0 | 83 | 0 |
| GT116 | 0 | 361 | 0 |

## PL

| Family | Sample Size | Population | Percentage in the test sets |
|---|---|---|---|
| PL0 | 2 | 2275 | 0.09 |
| PL1 | 63 | 13024 | 0.48 |
| PL2 | 4 | 856 | 0.47 |
| PL3 | 18 | 2579 | 0.7 |
| PL4 | 9 | 1278 | 0.7 |
| PL5 | 1 | 1589 | 0.06 |
| PL6 | 2 | 941 | 0.21 |
| PL7 | 11 | 3148 | 0.35 |
| PL8 | 4 | 4438 | 0.09 |
| PL9 | 14 | 3707 | 0.38 |
| PL10 | 7 | 1118 | 0.63 |
| PL11 | 13 | 2187 | 0.59 |
| PL12 | 2 | 2099 | 0.1 |
| PL13 | 0 | 120 | 0 |
| PL14 | 0 | 410 | 0 |
| PL15 | 1 | 483 | 0.21 |
| PL16 | 0 | 424 | 0 |
| PL17 | 2 | 1098 | 0.18 |
| PL18 | 0 | 47 | 0 |
| PL19 | 0 | 0 | 0 |
| PL20 | 5 | 81 | 6.17 |
| PL21 | 0 | 78 | 0 |
| PL22 | 8 | 2388 | 0.34 |
| PL23 | 0 | 342 | 0 |
| PL24 | 0 | 62 | 0 |
| PL25 | 0 | 56 | 0 |
| PL26 | 6 | 1015 | 0.59 |
| PL27 | 2 | 150 | 1.33 |
| PL28 | 0 | 30 | 0 |
| PL29 | 1 | 186 | 0.54 |
| PL30 | 0 | 33 | 0 |
| PL31 | 0 | 418 | 0 |
| PL32 | 0 | 15 | 0 |
| PL33 | 4 | 963 | 0.42 |
| PL34 | 0 | 20 | 0 |
| PL35 | 0 | 461 | 0 |
| PL36 | 0 | 72 | 0 |
| PL37 | 0 | 56 | 0 |
| PL38 | 9 | 1740 | 0.52 |
| PL39 | 0 | 26 | 0 |
| PL40 | 0 | 238 | 0 |
| PL41 | 0 | 41 | 0 |
| PL42 | 4 | 658 | 0.61 |

## CE

| Family | Sample Size | Population | Percentage in the test sets |
|---|---|---|---|
| CE0 | 31 | 3592 | 0.86 |
| CE1 | 22 | 5659 | 0.39 |
| CE2 | 5 | 958 | 0.52 |
| CE3 | 4 | 645 | 0.62 |
| CE4 | 95 | 44485 | 0.21 |
| CE5 | 40 | 4979 | 0.8 |
| CE6 | 3 | 648 | 0.46 |
| CE7 | 11 | 3447 | 0.32 |
| CE8 | 39 | 11325 | 0.34 |
| CE9 | 45 | 25515 | 0.18 |
| CE10 | 0 | 0 | 0 |
| CE11 | 22 | 17605 | 0.12 |
| CE12 | 33 | 4987 | 0.66 |
| CE13 | 0 | 552 | 0 |
| CE14 | 13 | 8293 | 0.16 |
| CE15 | 5 | 756 | 0.66 |
| CE16 | 16 | 241 | 6.64 |
| CE17 | 0 | 28 | 0 |
| CE18 | 1 | 47 | 2.13 |
| CE19 | 0 | 356 | 0 |
| CE20 | 14 | 4332 | 0.32 |

## AA

| Family | Sample Size | Population | Percentage in the test sets |
|---|---|---|---|
| AA0 | 5 | 139 | 3.6 |
| AA1 | 85 | 6951 | 1.22 |
| AA2 | 12 | 974 | 1.23 |
| AA3 | 75 | 3412 | 2.2 |
| AA4 | 3 | 88 | 3.41 |
| AA5 | 14 | 996 | 1.41 |
| AA6 | 36 | 874 | 4.12 |
| AA7 | 4 | 143 | 2.8 |
| AA8 | 5 | 223 | 2.24 |
| AA9 | 24 | 1160 | 2.07 |
| AA10 | 16 | 10184 | 0.16 |
| AA11 | 11 | 299 | 3.68 |
| AA12 | 7 | 124 | 5.65 |
| AA13 | 1 | 47 | 2.13 |
| AA14 | 4 | 67 | 5.97 |
| AA15 | 0 | 459 | 0 |
| AA16 | 5 | 88 | 5.68 |
| AA17 | 0 | 418 | 0 |
| AA18 | 0 | 0 | 0 |

## CBM

| Family | Sample Size | Population | Percentage in the test sets |
|---|---|---|---|
| CBM0 | 16 | 3541 | 0.45 |
| CBM1 | 45 | 2069 | 2.17 |
| CBM2 | 32 | 11256 | 0.28 |
| CBM3 | 35 | 2621 | 1.34 |
| CBM4 | 9 | 1274 | 0.71 |
| CBM5 | 18 | 12309 | 0.15 |
| CBM6 | 41 | 6383 | 0.64 |
| CBM7 | 0 | 0 | 0 |
| CBM8 | 0 | 128 | 0 |
| CBM9 | 11 | 958 | 1.15 |
| CBM10 | 8 | 336 | 2.38 |
| CBM11 | 0 | 209 | 0 |
| CBM12 | 7 | 3536 | 0.2 |
| CBM13 | 44 | 16215 | 0.27 |
| CBM14 | 2 | 3734 | 0.05 |
| CBM15 | 0 | 19 | 0 |
| CBM16 | 8 | 1442 | 0.55 |
| CBM17 | 1 | 63 | 1.59 |
| CBM18 | 52 | 2653 | 1.96 |
| CBM19 | 10 | 176 | 5.68 |
| CBM20 | 27 | 3223 | 0.84 |
| CBM21 | 31 | 1242 | 2.5 |
| CBM22 | 9 | 1670 | 0.54 |
| CBM23 | 0 | 252 | 0 |
| CBM24 | 9 | 125 | 7.2 |
| CBM25 | 1 | 550 | 0.18 |
| CBM26 | 3 | 832 | 0.36 |
| CBM27 | 0 | 130 | 0 |
| CBM28 | 1 | 72 | 1.39 |
| CBM29 | 0 | 1 | 0 |
| CBM30 | 2 | 39 | 5.13 |
| CBM31 | 0 | 19 | 0 |
| CBM32 | 39 | 13822 | 0.28 |
| CBM33 | 0 | 0 | 0 |
| CBM34 | 16 | 9026 | 0.18 |
| CBM35 | 38 | 5437 | 0.7 |
| CBM36 | 3 | 200 | 1.5 |
| CBM37 | 0 | 98 | 0 |
| CBM38 | 3 | 754 | 0.4 |
| CBM39 | 0 | 275 | 0 |
| CBM40 | 0 | 905 | 0 |
| CBM41 | 3 | 4818 | 0.06 |
| CBM42 | 9 | 1432 | 0.63 |
| CBM43 | 40 | 3095 | 1.29 |
| CBM44 | 0 | 34 | 0 |
| CBM45 | 3 | 198 | 1.52 |
| CBM46 | 3 | 334 | 0.9 |
| CBM47 | 1 | 1208 | 0.08 |
| CBM48 | 141 | 57970 | 0.24 |
| CBM49 | 0 | 210 | 0 |
| CBM50 | 157 | 112373 | 0.14 |
| CBM51 | 3 | 2582 | 0.12 |
| CBM52 | 1 | 24 | 4.17 |
| CBM53 | 7 | 180 | 3.89 |
| CBM54 | 3 | 404 | 0.74 |
| CBM55 | 0 | 19 | 0 |
| CBM56 | 1 | 360 | 0.28 |
| CBM57 | 1 | 1519 | 0.07 |
| CBM58 | 0 | 85 | 0 |
| CBM59 | 0 | 115 | 0 |
| CBM60 | 1 | 145 | 0.69 |
| CBM61 | 4 | 1047 | 0.38 |
| CBM62 | 1 | 534 | 0.19 |
| CBM63 | 6 | 1600 | 0.38 |
| CBM64 | 0 | 52 | 0 |
| CBM65 | 0 | 30 | 0 |
| CBM66 | 4 | 1233 | 0.32 |
| CBM67 | 3 | 1039 | 0.29 |
| CBM68 | 0 | 1093 | 0 |
| CBM69 | 0 | 411 | 0 |
| CBM70 | 0 | 539 | 0 |
| CBM71 | 0 | 238 | 0 |
| CBM72 | 0 | 68 | 0 |
| CBM73 | 9 | 4743 | 0.19 |
| CBM74 | 0 | 41 | 0 |
| CBM75 | 0 | 3 | 0 |
| CBM76 | 0 | 5 | 0 |
| CBM77 | 0 | 104 | 0 |
| CBM78 | 0 | 9 | 0 |
| CBM79 | 0 | 16 | 0 |
| CBM80 | 0 | 4 | 0 |
| CBM81 | 0 | 7 | 0 |
| CBM82 | 0 | 12 | 0 |
| CBM83 | 0 | 6 | 0 |
| CBM84 | 0 | 63 | 0 |
| CBM85 | 0 | 153 | 0 |
| CBM86 | 0 | 11 | 0 |
| CBM87 | 1 | 61 | 1.64 |
| CBM88 | 0 | 59 | 0 |
| CBM89 | 0 | 16 | 0 |
| CBM90 | 0 | 22 | 0 |
| CBM91 | 51 | 11347 | 0.45 |
| CBM92 | 1 | 882 | 0.11 |
| CBM93 | 0 | 276 | 0 |
| CBM94 | 0 | 294 | 0 |
| CBM95 | 0 | 130 | 0 |
| CBM96 | 2 | 490 | 0.41 |
| CBM97 | 1 | 202 | 0.5 |

# 2 Evaluation of CAZyme/non-CAZyme classification

## 2.1 Summary of CAZyme/non-CAZyme classification

### SI Table 3: Summary of binary CAZyme/non-CAZyme classification (overleaf)

SI table 3 lists the statistical parameter values for the binary CAZyme/non-CAZyme classification of protein sequences, reporting the mean, standard deviation,and lower and upper 95% confidence intervals (CIs) across all test sets for each statistical parameter.

| Classifier | | CUPP | dbCAN_2 | dbCAN_2:DIAMOND | dbCAN_2:HMMER | dbCAN_2:Hotpep | dbCAN_3 | dbCAN_3:DIAMOND | dbCAN_3:eCAMI | dbCAN_3:HMMER | dbCAN_4 | dbCAN_4:dbCAN-sub | dbCAN_4:DIAMOND | dbCAN_4:HMMER |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **Specificity** | **Mean** | 0.982 | 0.977 | 0.974 | 0.979 | 0.976 | 0.959 | 0.967 | 0.975 | 0.979 | 0.962 | 0.978 | 0.967 | 0.979 |
| | **Standard Deviation** | 0.049 | 0.053 | 0.054 | 0.052 | 0.046 | 0.077 | 0.074 | 0.048 | 0.052 | 0.076 | 0.053 | 0.074 | 0.052 |
| | **Lower CI** | 0.971 | 0.966 | 0.962 | 0.967 | 0.966 | 0.942 | 0.950 | 0.965 | 0.967 | 0.945 | 0.966 | 0.950 | 0.967 |
| | **Upper CI** | 0.993 | 0.989 | 0.986 | 0.990 | 0.986 | 0.976 | 0.983 | 0.986 | 0.990 | 0.979 | 0.989 | 0.983 | 0.990 |
| **Sensitivity** | **Mean** | 0.854 | 0.901 | 0.916 | 0.878 | 0.808 | 0.987 | 0.969 | 0.852 | 0.888 | 0.984 | 0.950 | 0.969 | 0.890 |
| | **Standard Deviation** | 0.074 | 0.108 | 0.129 | 0.081 | 0.127 | 0.044 | 0.082 | 0.128 | 0.082 | 0.064 | 0.073 | 0.084 | 0.082 |
| | **Lower CI** | 0.837 | 0.877 | 0.887 | 0.860 | 0.780 | 0.977 | 0.951 | 0.823 | 0.870 | 0.970 | 0.934 | 0.951 | 0.872 |
| | **Upper CI** | 0.870 | 0.925 | 0.944 | 0.896 | 0.837 | 0.997 | 0.987 | 0.880 | 0.906 | 0.998 | 0.966 | 0.988 | 0.908 |
| **Precision** | **Mean** | 0.981 | 0.978 | 0.976 | 0.979 | 0.972 | 0.964 | 0.971 | 0.974 | 0.979 | 0.967 | 0.979 | 0.971 | 0.979 |
| | **Standard Deviation** | 0.045 | 0.045 | 0.045 | 0.045 | 0.053 | 0.056 | 0.054 | 0.046 | 0.045 | 0.055 | 0.043 | 0.055 | 0.045 |
| | **Lower CI** | 0.971 | 0.968 | 0.966 | 0.969 | 0.960 | 0.952 | 0.959 | 0.964 | 0.969 | 0.955 | 0.970 | 0.959 | 0.969 |
| | **Upper CI** | 0.991 | 0.988 | 0.986 | 0.989 | 0.984 | 0.977 | 0.983 | 0.984 | 0.989 | 0.980 | 0.989 | 0.983 | 0.989 |
| **F1-score Mean** | **Mean** | 0.911 | 0.933 | 0.938 | 0.923 | 0.877 | 0.974 | 0.966 | 0.903 | 0.928 | 0.973 | 0.962 | 0.967 | 0.929 |
| | **Standard Deviation** | 0.052 | 0.079 | 0.092 | 0.064 | 0.092 | 0.041 | 0.066 | 0.087 | 0.065 | 0.053 | 0.056 | 0.069 | 0.065 |
| | **Lower CI** | 0.899 | 0.916 | 0.917 | 0.908 | 0.857 | 0.965 | 0.952 | 0.883 | 0.914 | 0.961 | 0.949 | 0.951 | 0.915 |
| | **Upper CI** | 0.923 | 0.951 | 0.959 | 0.937 | 0.897 | 0.983 | 0.981 | 0.922 | 0.943 | 0.985 | 0.974 | 0.982 | 0.944 |
| **Accuracy** | **Mean** | 0.918 | 0.939 | 0.945 | 0.928 | 0.892 | 0.973 | 0.968 | 0.913 | 0.933 | 0.973 | 0.964 | 0.968 | 0.934 |
| | **Standard Deviation** | 0.046 | 0.060 | 0.067 | 0.047 | 0.071 | 0.044 | 0.054 | 0.068 | 0.048 | 0.049 | 0.045 | 0.055 | 0.048 |
| | **Lower CI** | 0.908 | 0.926 | 0.930 | 0.918 | 0.876 | 0.963 | 0.956 | 0.898 | 0.923 | 0.962 | 0.954 | 0.956 | 0.924 |
| | **Upper CI** | 0.928 | 0.953 | 0.960 | 0.939 | 0.908 | 0.983 | 0.980 | 0.929 | 0.944 | 0.984 | 0.974 | 0.980 | 0.945 |

## 2.2 Output of testing for statistically significant difference in performance between the tools

### SI Figure 1: Evaluation of binary CAZyme/non-CAZyme classification

SI figure **??** plots the value of each statistical parameter for each test, per CAZyme classifier.

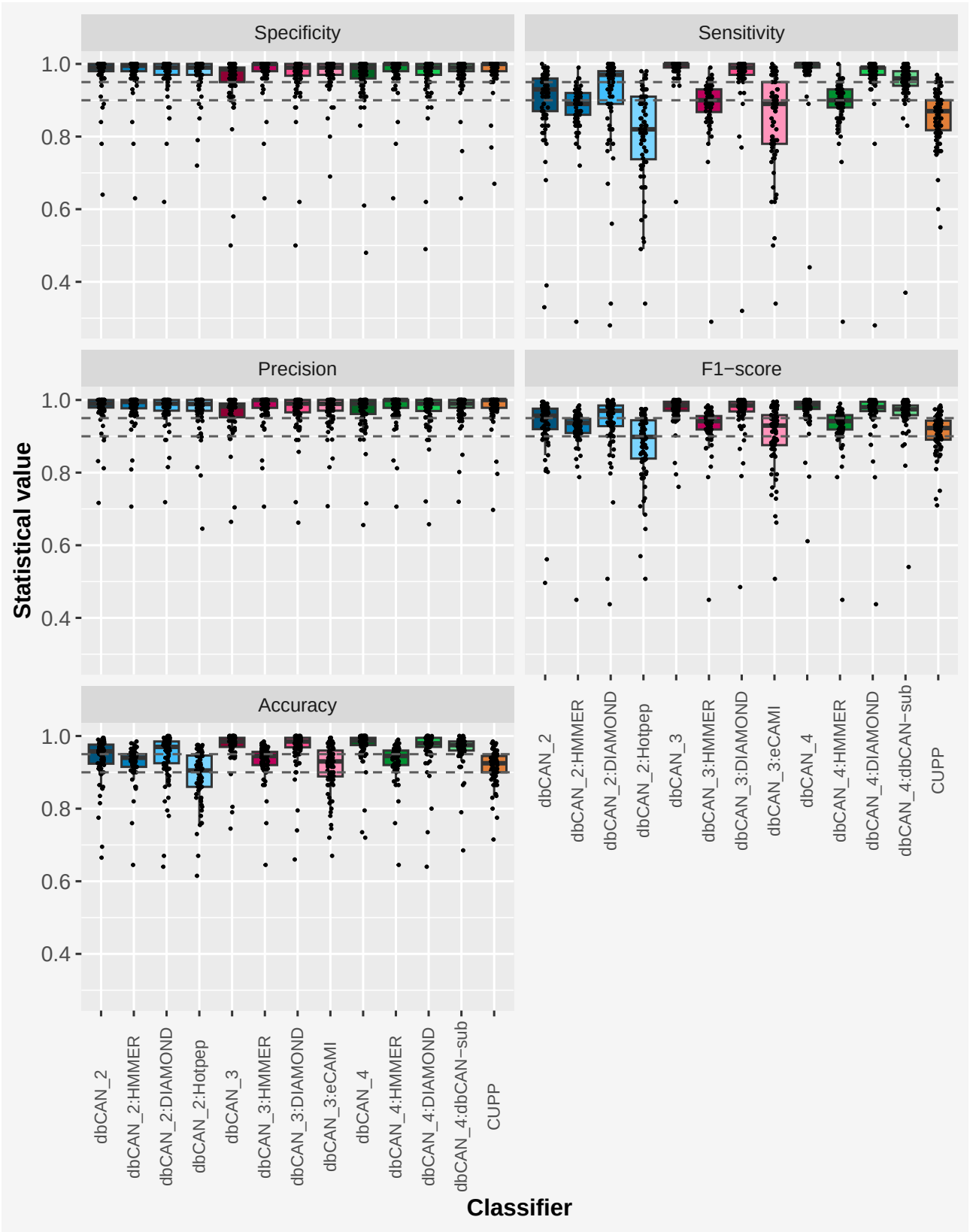
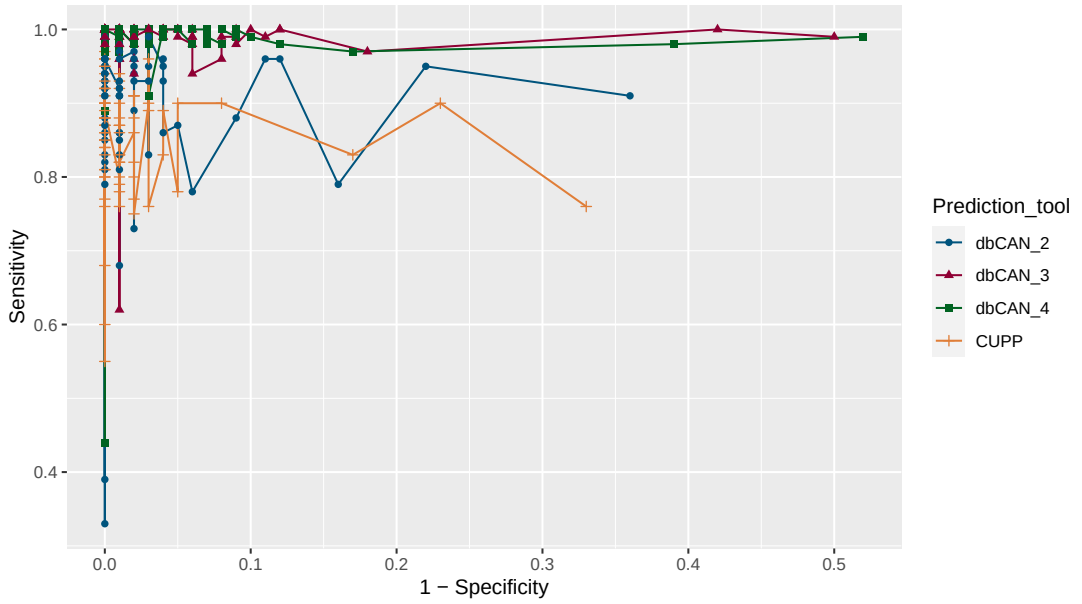



Figure 1: One-dimensional scatter plot overlaying a box and whisker plot, where each scatter plot point represents a test set and the corresponding statistical parameter value. Dashed lines indicate statistical values of 0.9 and 0.95.

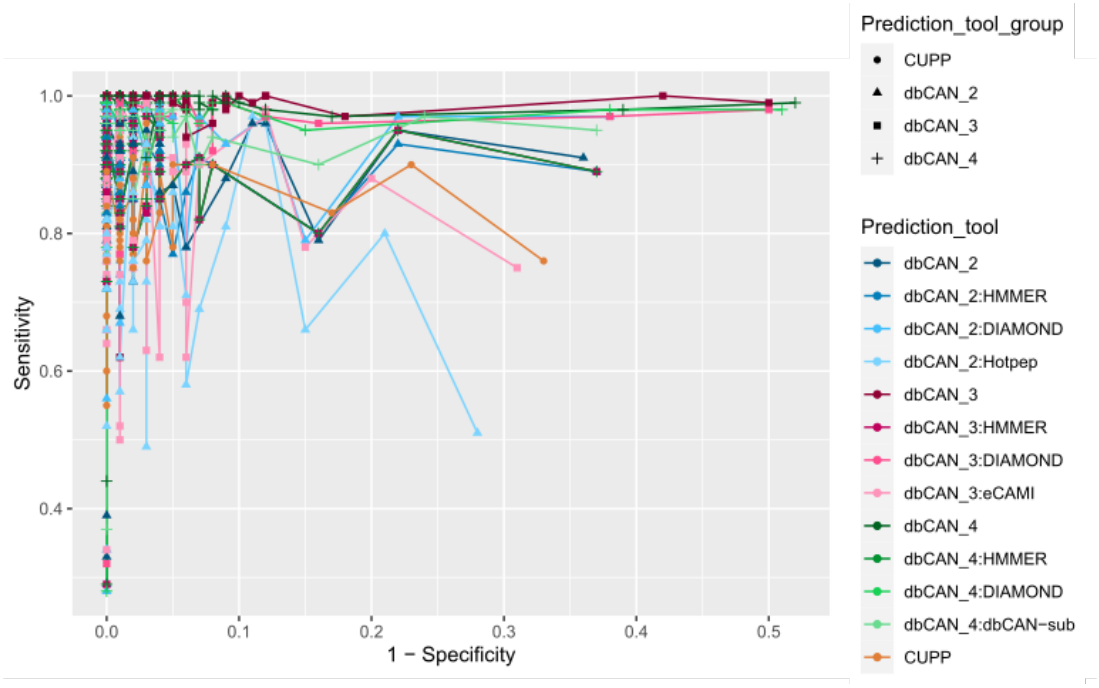
??

## SI Figure 2: ROC curve of binary CAZyme/non-CAZyme classification

SI Figure 2 plots the Receiver Operator Characteristic (ROC) curve, specifically plotting sensitivity against 1-specificity of the binary CAZyme/non-CAZyme classification for each test set.




(a) ROC main classifiers




(b) ROC all classifiers

Figure 2: Receiver Operator Characteristic (ROC) curve of the binary CAZyme/non-CAZyme classification, across all test sets per CAZyme classifier plotting [A] only the main classifiers and [B] all classifiers and tools incorporated into each classifier.

## SI Table 4: Tukey HSD test to measure the statistically significant difference between the mean F1-score for CAZyme/non-CAZyme classification (overleaf)

Tukey HSD test of the mean F1-score between the main classifiers (CUPP, dbCAN_2, dbCAN_3, and dbCAN_4) and incorporated tools, reporting the lower and upper 95% confidence interval, and the adjusted p-value (P-Adj).

| Classifier 1 - vs - Classifier 2 | Mean Difference | Lower Confidence Interval | Upper Confidence Interval | Adjusted P-value |
|---|---|---|---|---|
| dbCAN_2-CUPP | 0.022283157 | -0.014177588 | 0.058743902 | 0.713229893 |
| dbCAN_2:DIAMOND-CUPP | 0.027081249 | -0.009379497 | 0.063541994 | 0.398221417 |
| dbCAN_2:HMMER-CUPP | 0.011557243 | -0.024903502 | 0.048017988 | 0.997946457 |
| dbCAN_2:Hotpep-CUPP | -0.033870876 | -0.070331621 | 0.002589869 | 0.099398373 |
| dbCAN_3-CUPP | 0.063226624 | 0.026765879 | 0.099687369 | 8.67E-07 |
| dbCAN_3:DIAMOND-CUPP | 0.055435558 | 0.018974813 | 0.091896303 | 3.98E-05 |
| dbCAN_3:eCAMI-CUPP | -0.008356874 | -0.044817619 | 0.028103871 | 0.999924847 |
| dbCAN_3:HMMER-CUPP | 0.017288455 | -0.019172291 | 0.0537492 | 0.937468195 |
| dbCAN_4-CUPP | 0.062315777 | 0.025855032 | 0.098776522 | 1.39E-06 |
| dbCAN_4:dbCAN-sub-CUPP | 0.050883313 | 0.014422568 | 0.087344058 | 0.000297288 |
| dbCAN_4:DIAMOND-CUPP | 0.055632677 | 0.019171932 | 0.092093422 | 3.63E-05 |
| dbCAN_4:HMMER-CUPP | 0.018405523 | -0.018055222 | 0.054866268 | 0.903910193 |
| dbCAN_2:DIAMOND-dbCAN_2 | 0.004798091 | -0.031662654 | 0.041258837 | 0.99999985 |
| dbCAN_2:HMMER-dbCAN_2 | -0.010725914 | -0.04718666 | 0.025734831 | 0.999008914 |
| dbCAN_2:Hotpep-dbCAN_2 | -0.056154033 | -0.092614778 | -0.019693288 | 2.85E-05 |
| dbCAN_3-dbCAN_2 | 0.040943467 | 0.004482722 | 0.077404212 | 0.012698109 |
| dbCAN_3:DIAMOND-dbCAN_2 | 0.033152401 | -0.003308345 | 0.069613146 | 0.118529315 |
| dbCAN_3:eCAMI-dbCAN_2 | -0.030640031 | -0.067100776 | 0.005820714 | 0.20811152 |
| dbCAN_3:HMMER-dbCAN_2 | -0.004994703 | -0.041455448 | 0.031466043 | 0.999999762 |
| dbCAN_4-dbCAN_2 | 0.04003262 | 0.003571875 | 0.076493365 | 0.017073759 |
| dbCAN_4:dbCAN-sub-dbCAN_2 | 0.028600156 | -0.007860589 | 0.065060901 | 0.308609131 |
| dbCAN_4:DIAMOND-dbCAN_2 | 0.03334952 | -0.003111225 | 0.069810265 | 0.113014504 |
| dbCAN_4:HMMER-dbCAN_2 | -0.003877634 | -0.040338379 | 0.032583111 | 0.999999987 |
| dbCAN_2:HMMER-dbCAN_2:DIAMOND | -0.015524006 | -0.051984751 | 0.020936739 | 0.972211635 |
| dbCAN_2:Hotpep-dbCAN_2:DIAMOND | -0.060952125 | -0.09741287 | -0.02449138 | 2.78E-06 |
| dbCAN_3-dbCAN_2:DIAMOND | 0.036145376 | -0.000315369 | 0.072606121 | 0.054596755 |
| dbCAN_3:DIAMOND-dbCAN_2:DIAMOND | 0.028354309 | -0.008106436 | 0.064815054 | 0.322336993 |
| dbCAN_3:eCAMI-dbCAN_2:DIAMOND | -0.035438122 | -0.071898867 | 0.001022623 | 0.066218667 |
| dbCAN_3:HMMER-dbCAN_2:DIAMOND | -0.009792794 | -0.046253539 | 0.026667951 | 0.999604023 |
| dbCAN_4-dbCAN_2:DIAMOND | 0.035234529 | -0.001226216 | 0.071695274 | 0.069924332 |
| dbCAN_4:dbCAN-sub-dbCAN_2:DIAMOND | 0.023802065 | -0.01265868 | 0.06026281 | 0.615149117 |
| dbCAN_4:DIAMOND-dbCAN_2:DIAMOND | 0.028551428 | -0.007909317 | 0.065012173 | 0.311304174 |
| dbCAN_4:HMMER-dbCAN_2:DIAMOND | -0.008675725 | -0.04513647 | 0.02778502 | 0.999888016 |
| dbCAN_2:Hotpep-dbCAN_2:HMMER | -0.045428119 | -0.081888864 | -0.008967374 | 0.002616469 |
| dbCAN_3-dbCAN_2:HMMER | 0.051669382 | 0.015208636 | 0.088130127 | 0.000212685 |
| dbCAN_3:DIAMOND-dbCAN_2:HMMER | 0.043878315 | 0.00741757 | 0.08033906 | 0.004617979 |
| dbCAN_3:eCAMI-dbCAN_2:HMMER | -0.019914116 | -0.056374861 | 0.016546629 | 0.842812921 |
| dbCAN_3:HMMER-dbCAN_2:HMMER | 0.005731212 | -0.030729533 | 0.042191957 | 0.999998848 |
| dbCAN_4-dbCAN_2:HMMER | 0.050758535 | 0.014297789 | 0.08721928 | 0.000313367 |
| dbCAN_4:dbCAN-sub-dbCAN_2:HMMER | 0.03932607 | 0.002865325 | 0.075786816 | 0.021352938 |
| dbCAN_4:DIAMOND-dbCAN_2:HMMER | 0.044075434 | 0.007614689 | 0.080536179 | 0.004301534 |
| dbCAN_4:HMMER-dbCAN_2:HMMER | 0.006848281 | -0.029612464 | 0.043309026 | 0.999991454 |
| dbCAN_3-dbCAN_2:Hotpep | 0.0970975 | 0.060636755 | 0.133558245 | 3.99E-13 |
| dbCAN_3:DIAMOND-dbCAN_2:Hotpep | 0.089306434 | 0.052845689 | 0.125767179 | 5.34E-13 |
| dbCAN_3:eCAMI-dbCAN_2:Hotpep | 0.025514002 | -0.010946743 | 0.061974748 | 0.500034421 |
| dbCAN_3:HMMER-dbCAN_2:Hotpep | 0.051159331 | 0.014698586 | 0.087620076 | 0.000264465 |
| dbCAN_4-dbCAN_2:Hotpep | 0.096186653 | 0.059725908 | 0.132647398 | 3.98E-13 |
| dbCAN_4:dbCAN-sub-dbCAN_2:Hotpep | 0.084754189 | 0.048293444 | 0.121214934 | 2.56E-12 |
| dbCAN_4:DIAMOND-dbCAN_2:Hotpep | 0.089503553 | 0.053042808 | 0.125964298 | 5.20E-13 |
| dbCAN_4:HMMER-dbCAN_2:Hotpep | 0.052276399 | 0.015815654 | 0.088737145 | 0.000163634 |
| dbCAN_3:DIAMOND-dbCAN_3 | -0.007791067 | -0.044251812 | 0.028669678 | 0.999964712 |
| dbCAN_3:eCAMI-dbCAN_3 | -0.071583498 | -0.108044243 | -0.035122753 | 8.60E-09 |
| dbCAN_3:HMMER-dbCAN_3 | -0.04593817 | -0.082398915 | -0.009477425 | 0.002159525 |
| dbCAN_4-dbCAN_3 | -0.000910847 | -0.037371592 | 0.035549898 | 1 |
| dbCAN_4:dbCAN-sub-dbCAN_3 | -0.012343311 | -0.048804056 | 0.024117434 | 0.996175862 |
| dbCAN_4:DIAMOND-dbCAN_3 | -0.007593947 | -0.044054693 | 0.028866798 | 0.999973306 |
| dbCAN_4:HMMER-dbCAN_3 | -0.044821101 | -0.081281846 | -0.008360356 | 0.003277467 |
| dbCAN_3:eCAMI-dbCAN_3:DIAMOND | -0.063792431 | -0.100253176 | -0.027331686 | 6.45E-07 |
| dbCAN_3:HMMER-dbCAN_3:DIAMOND | -0.038147103 | -0.074607848 | -0.001686358 | 0.030641611 |
| dbCAN_4-dbCAN_3:DIAMOND | 0.00688022 | -0.029580525 | 0.043340965 | 0.999991001 |
| dbCAN_4:dbCAN-sub-dbCAN_3:DIAMOND | -0.004552244 | -0.04101299 | 0.031908501 | 0.999999918 |
| dbCAN_4:DIAMOND-dbCAN_3:DIAMOND | 0.000197119 | -0.036263626 | 0.036657864 | 1 |
| dbCAN_4:HMMER-dbCAN_3:DIAMOND | -0.037030034 | -0.073490779 | -0.000569289 | 0.042535486 |
| dbCAN_3:HMMER-dbCAN_3:eCAMI | 0.025645328 | -0.010815417 | 0.062106073 | 0.491269354 |
| dbCAN_4-dbCAN_3:eCAMI | 0.070672651 | 0.034211906 | 0.107133396 | 1.46E-08 |
| dbCAN_4:dbCAN-sub-dbCAN_3:eCAMI | 0.059240187 | 0.022779442 | 0.095700932 | 6.52E-06 |
| dbCAN_4:DIAMOND-dbCAN_3:eCAMI | 0.06398955 | 0.027528805 | 0.100450296 | 5.82E-07 |
| dbCAN_4:HMMER-dbCAN_3:eCAMI | 0.026762397 | -0.009698348 | 0.063223142 | 0.418325189 |
| dbCAN_4-dbCAN_3:HMMER | 0.045027323 | 0.008566578 | 0.081488068 | 0.003037208 |
| dbCAN_4:dbCAN-sub-dbCAN_3:HMMER | 0.033594859 | -0.002865886 | 0.070055604 | 0.106434458 |
| dbCAN_4:DIAMOND-dbCAN_3:HMMER | 0.038344222 | 0.001883477 | 0.074804967 | 0.028876665 |
| dbCAN_4:HMMER-dbCAN_3:HMMER | 0.001117069 | -0.035343676 | 0.037577814 | 1 |
| dbCAN_4:dbCAN-sub-dbCAN_4 | -0.011432464 | -0.047893209 | 0.025028281 | 0.998150174 |
| dbCAN_4:DIAMOND-dbCAN_4 | -0.0066831 | -0.043143846 | 0.029777645 | 0.999993488 |
| dbCAN_4:HMMER-dbCAN_4 | -0.043910254 | -0.080370999 | -0.007449509 | 0.004565285 |
| dbCAN_4:DIAMOND-dbCAN_4:dbCAN-sub | 0.004749364 | -0.031711381 | 0.041210109 | 0.999999867 |
| dbCAN_4:HMMER-dbCAN_4:dbCAN-sub | -0.03247779 | -0.068938535 | 0.003982955 | 0.138996641 |
| dbCAN_4:HMMER-dbCAN_4:DIAMOND | -0.037227154 | -0.073687899 | -0.000766408 | 0.04018476 |

## SI Table 5: Tukey HSD test to measure the statistically significant difference between the mean sensitivity for CAZyme/non-CAZyme classification (overleaf)

Tukey HSD test of the mean sensitivity between the main classifiers (CUPP, dbCAN_2, dbCAN_3, and dbCAN_4) and incorporated tools, reporting the lower and upper 95% confidence interval, and the adjusted p-value (P-Adj).

| Classifier 1 - vs - Classifier 2 | Mean Difference | Lower Confidence Interval | Upper Confidence Interval | Adjusted P-value |
|---|---|---|---|---|
| dbCAN_2-CUPP | 0.047375 | -0.001178458 | 0.095928458 | 0.063806877 |
| dbCAN_2:DIAMOND-CUPP | 0.061875 | 0.013321542 | 0.110428458 | 0.001767064 |
| dbCAN_2:HMMER-CUPP | 0.024 | -0.024553458 | 0.072553458 | 0.916514145 |
| dbCAN_2:Hotpep-CUPP | -0.0455 | -0.094053458 | 0.003053458 | 0.092250165 |
| dbCAN_3-CUPP | 0.13325 | 0.084696542 | 0.181803458 | 4.09E-13 |
| dbCAN_3:DIAMOND-CUPP | 0.115 | 0.066446542 | 0.163553458 | 1.15E-12 |
| dbCAN_3:eCAMI-CUPP | -0.00225 | -0.050803458 | 0.046303458 | 1 |
| dbCAN_3:HMMER-CUPP | 0.034125 | -0.014428458 | 0.082678458 | 0.492568515 |
| dbCAN_4-CUPP | 0.13025 | 0.081696542 | 0.178803458 | 4.03E-13 |
| dbCAN_4:dbCAN-sub-CUPP | 0.095875 | 0.047321542 | 0.144428458 | 6.76E-09 |
| dbCAN_4:DIAMOND-CUPP | 0.1155 | 0.066946542 | 0.164053458 | 9.92E-13 |
| dbCAN_4:HMMER-CUPP | 0.036125 | -0.012428458 | 0.084678458 | 0.39532421 |
| dbCAN_2:DIAMOND-dbCAN_2 | 0.0145 | -0.034053458 | 0.063053458 | 0.998849919 |
| dbCAN_2:HMMER-dbCAN_2 | -0.023375 | -0.071928458 | 0.025178458 | 0.930386981 |
| dbCAN_2:Hotpep-dbCAN_2 | -0.092875 | -0.141428458 | -0.044321542 | 2.48E-08 |
| dbCAN_3-dbCAN_2 | 0.085875 | 0.037321542 | 0.134428458 | 4.47E-07 |
| dbCAN_3:DIAMOND-dbCAN_2 | 0.067625 | 0.019071542 | 0.116178458 | 0.000310238 |
| dbCAN_3:eCAMI-dbCAN_2 | -0.049625 | -0.098178458 | -0.001071542 | 0.039741787 |
| dbCAN_3:HMMER-dbCAN_2 | -0.01325 | -0.061803458 | 0.035303458 | 0.999534102 |
| dbCAN_4-dbCAN_2 | 0.082875 | 0.034321542 | 0.131428458 | 1.45E-06 |
| dbCAN_4:dbCAN-sub-dbCAN_2 | 0.0485 | -5.35E-05 | 0.097053458 | 0.050566169 |
| dbCAN_4:DIAMOND-dbCAN_2 | 0.068125 | 0.019571542 | 0.116678458 | 0.000264636 |
| dbCAN_4:HMMER-dbCAN_2 | -0.01125 | -0.059803458 | 0.037303458 | 0.999915605 |
| dbCAN_2:HMMER-dbCAN_2:DIAMOND | -0.037875 | -0.086428458 | 0.010678458 | 0.317411004 |
| dbCAN_2:Hotpep-dbCAN_2:DIAMOND | -0.107375 | -0.155928458 | -0.058821542 | 3.34E-11 |
| dbCAN_3-dbCAN_2:DIAMOND | 0.071375 | 0.022821542 | 0.119928458 | 9.15E-05 |
| dbCAN_3:DIAMOND-dbCAN_2:DIAMOND | 0.053125 | 0.004571542 | 0.101678458 | 0.017848783 |
| dbCAN_3:eCAMI-dbCAN_2:DIAMOND | -0.064125 | -0.112678458 | -0.015571542 | 0.000912355 |
| dbCAN_3:HMMER-dbCAN_2:DIAMOND | -0.02775 | -0.076303458 | 0.020803458 | 0.796650206 |
| dbCAN_4-dbCAN_2:DIAMOND | 0.068375 | 0.019821542 | 0.116928458 | 0.000244305 |
| dbCAN_4:dbCAN-sub-dbCAN_2:DIAMOND | 0.034 | -0.014553458 | 0.082553458 | 0.498833688 |
| dbCAN_4:DIAMOND-dbCAN_2:DIAMOND | 0.053625 | 0.005071542 | 0.102178458 | 0.015823284 |
| dbCAN_4:HMMER-dbCAN_2:DIAMOND | -0.02575 | -0.074303458 | 0.022803458 | 0.868346985 |
| dbCAN_2:Hotpep-dbCAN_2:HMMER | -0.0695 | -0.118053458 | -0.020946542 | 0.000169862 |
| dbCAN_3-dbCAN_2:HMMER | 0.10925 | 0.060696542 | 0.157803458 | 1.35E-11 |
| dbCAN_3:DIAMOND-dbCAN_2:HMMER | 0.091 | 0.042446542 | 0.139553458 | 5.49E-08 |
| dbCAN_3:eCAMI-dbCAN_2:HMMER | -0.02625 | -0.074803458 | 0.022303458 | 0.85204086 |
| dbCAN_3:HMMER-dbCAN_2:HMMER | 0.010125 | -0.038428458 | 0.058678458 | 0.999972946 |
| dbCAN_4-dbCAN_2:HMMER | 0.10625 | 0.057696542 | 0.154803458 | 5.72E-11 |
| dbCAN_4:dbCAN-sub-dbCAN_2:HMMER | 0.071875 | 0.023321542 | 0.120428458 | 7.73E-05 |
| dbCAN_4:DIAMOND-dbCAN_2:HMMER | 0.0915 | 0.042946542 | 0.140053458 | 4.45E-08 |
| dbCAN_4:HMMER-dbCAN_2:HMMER | 0.012125 | -0.036428458 | 0.060678458 | 0.999813699 |
| dbCAN_3-dbCAN_2:Hotpep | 0.17875 | 0.130196542 | 0.227303458 | 2.86E-13 |
| dbCAN_3:DIAMOND-dbCAN_2:Hotpep | 0.1605 | 0.111946542 | 0.209053458 | 3.68E-13 |
| dbCAN_3:eCAMI-dbCAN_2:Hotpep | 0.04325 | -0.005303458 | 0.091803458 | 0.138984543 |
| dbCAN_3:HMMER-dbCAN_2:Hotpep | 0.079625 | 0.031071542 | 0.128178458 | 4.96E-06 |
| dbCAN_4-dbCAN_2:Hotpep | 0.17575 | 0.127196542 | 0.224303458 | 2.86E-13 |
| dbCAN_4:dbCAN-sub-dbCAN_2:Hotpep | 0.141375 | 0.092821542 | 0.189928458 | 4.08E-13 |
| dbCAN_4:DIAMOND-dbCAN_2:Hotpep | 0.161 | 0.112446542 | 0.209553458 | 3.65E-13 |
| dbCAN_4:HMMER-dbCAN_2:Hotpep | 0.081625 | 0.033071542 | 0.130178458 | 2.34E-06 |
| dbCAN_3:DIAMOND-dbCAN_3 | -0.01825 | -0.066803458 | 0.030303458 | 0.990175894 |
| dbCAN_3:eCAMI-dbCAN_3 | -0.1355 | -0.184053458 | -0.086946542 | 4.20E-13 |
| dbCAN_3:HMMER-dbCAN_3 | -0.099125 | -0.147678458 | -0.050571542 | 1.58E-09 |
| dbCAN_4-dbCAN_3 | -0.003 | -0.051553458 | 0.045553458 | 1 |
| dbCAN_4:dbCAN-sub-dbCAN_3 | -0.037375 | -0.085928458 | 0.011178458 | 0.338818388 |
| dbCAN_4:DIAMOND-dbCAN_3 | -0.01775 | -0.066303458 | 0.030803458 | 0.992308367 |
| dbCAN_4:HMMER-dbCAN_3 | -0.097125 | -0.145678458 | -0.048571542 | 3.89E-09 |
| dbCAN_3:eCAMI-dbCAN_3:DIAMOND | -0.11725 | -0.165803458 | -0.068696542 | 6.60E-13 |
| dbCAN_3:HMMER-dbCAN_3:DIAMOND | -0.080875 | -0.129428458 | -0.032321542 | 3.11E-06 |
| dbCAN_4-dbCAN_3:DIAMOND | 0.01525 | -0.033303458 | 0.063803458 | 0.998119718 |
| dbCAN_4:dbCAN-sub-dbCAN_3:DIAMOND | -0.019125 | -0.067678458 | 0.029428458 | 0.98531139 |
| dbCAN_4:DIAMOND-dbCAN_3:DIAMOND | 0.0005 | -0.048053458 | 0.049053458 | 1 |
| dbCAN_4:HMMER-dbCAN_3:DIAMOND | -0.078875 | -0.127428458 | -0.030321542 | 6.55E-06 |
| dbCAN_3:HMMER-dbCAN_3:eCAMI | 0.036375 | -0.012178458 | 0.084928458 | 0.38370878 |
| dbCAN_4-dbCAN_3:eCAMI | 0.1325 | 0.083946542 | 0.181053458 | 4.08E-13 |
| dbCAN_4:dbCAN-sub-dbCAN_3:eCAMI | 0.098125 | 0.049571542 | 0.146678458 | 2.49E-09 |
| dbCAN_4:DIAMOND-dbCAN_3:eCAMI | 0.11775 | 0.069196542 | 0.166303458 | 6.15E-13 |
| dbCAN_4:HMMER-dbCAN_3:eCAMI | 0.038375 | -0.010178458 | 0.086928458 | 0.296756154 |
| dbCAN_4-dbCAN_3:HMMER | 0.096125 | 0.047571542 | 0.144678458 | 6.05E-09 |
| dbCAN_4:dbCAN-sub-dbCAN_3:HMMER | 0.06175 | 0.013196542 | 0.110303458 | 0.001831768 |
| dbCAN_4:DIAMOND-dbCAN_3:HMMER | 0.081375 | 0.032821542 | 0.129928458 | 2.57E-06 |
| dbCAN_4:HMMER-dbCAN_3:HMMER | 0.002 | -0.046553458 | 0.050553458 | 1 |
| dbCAN_4:dbCAN-sub-dbCAN_4 | -0.034375 | -0.082928458 | 0.014178458 | 0.480085305 |
| dbCAN_4:DIAMOND-dbCAN_4 | -0.01475 | -0.063303458 | 0.033803458 | 0.998639772 |
| dbCAN_4:HMMER-dbCAN_4 | -0.094125 | -0.142678458 | -0.045571542 | 1.45E-08 |
| dbCAN_4:DIAMOND-dbCAN_4:dbCAN-sub | 0.019625 | -0.028928458 | 0.068178458 | 0.981772611 |
| dbCAN_4:HMMER-dbCAN_4:dbCAN-sub | -0.05975 | -0.108303458 | -0.011196542 | 0.003220516 |
| dbCAN_4:HMMER-dbCAN_4:DIAMOND | -0.079375 | -0.127928458 | -0.030821542 | 5.44E-06 |

## SI Table 6: Tukey HSD test to measure the statistically significant difference between the mean accuracy for CAZyme/non-CAZyme classification (overleaf)

Tukey HSD test of the mean accuracy between the main classifiers (CUPP, dbCAN_2, dbCAN_3, and dbCAN_4) and incorporated tools, reporting the lower and upper 95% confidence interval, and the adjusted p-value (P-Adj).

| Classifier 1 - vs - Classifier 2 | Mean Difference | Lower Confidence Interval | Upper Confidence Interval | Adjusted P-value |
|---|---|---|---|---|
| dbCAN_2-CUPP | 0.0214375 | -0.007352357 | 0.050227357 | 0.393975467 |
| dbCAN_2:DIAMOND-CUPP | 0.02725 | -0.001539857 | 0.056039857 | 0.084518677 |
| dbCAN_2:HMMER-CUPP | 0.0104375 | -0.018352357 | 0.039227357 | 0.99285914 |
| dbCAN_2:Hotpep-CUPP | -0.0256875 | -0.054477357 | 0.003102357 | 0.137265496 |
| dbCAN_3-CUPP | 0.055375 | 0.026585143 | 0.084164857 | 1.99E-08 |
| dbCAN_3:DIAMOND-CUPP | 0.049875 | 0.021085143 | 0.078664857 | 8.96E-07 |
| dbCAN_3:eCAMI-CUPP | -0.004375 | -0.033164857 | 0.024414857 | 0.999999216 |
| dbCAN_3:HMMER-CUPP | 0.0155625 | -0.013227357 | 0.044352357 | 0.852182049 |
| dbCAN_4-CUPP | 0.055375 | 0.026585143 | 0.084164857 | 1.99E-08 |
| dbCAN_4:dbCAN-sub-CUPP | 0.0458125 | 0.017022643 | 0.074602357 | 1.18E-05 |
| dbCAN_4:DIAMOND-CUPP | 0.05025 | 0.021460143 | 0.079039857 | 6.99E-07 |
| dbCAN_4:HMMER-CUPP | 0.0165625 | -0.012227357 | 0.045352357 | 0.789296176 |
| dbCAN_2:DIAMOND-dbCAN_2 | 0.0058125 | -0.022977357 | 0.034602357 | 0.999981017 |
| dbCAN_2:HMMER-dbCAN_2 | -0.011 | -0.039789857 | 0.017789857 | 0.988675544 |
| dbCAN_2:Hotpep-dbCAN_2 | -0.047125 | -0.075914857 | -0.018335143 | 5.25E-06 |
| dbCAN_3-dbCAN_2 | 0.0339375 | 0.005147643 | 0.062727357 | 0.006351466 |
| dbCAN_3:DIAMOND-dbCAN_2 | 0.0284375 | -0.000352357 | 0.057227357 | 0.056606885 |
| dbCAN_3:eCAMI-dbCAN_2 | -0.0258125 | -0.054602357 | 0.002977357 | 0.132284366 |
| dbCAN_3:HMMER-dbCAN_2 | -0.005875 | -0.034664857 | 0.022914857 | 0.999978653 |
| dbCAN_4-dbCAN_2 | 0.0339375 | 0.005147643 | 0.062727357 | 0.006351466 |
| dbCAN_4:dbCAN-sub-dbCAN_2 | 0.024375 | -0.004414857 | 0.053164857 | 0.198369746 |
| dbCAN_4:DIAMOND-dbCAN_2 | 0.0288125 | 2.26E-05 | 0.057602357 | 0.049598878 |
| dbCAN_4:HMMER-dbCAN_2 | -0.004875 | -0.033664857 | 0.023914857 | 0.999997323 |
| dbCAN_2:HMMER-dbCAN_2:DIAMOND | -0.0168125 | -0.045602357 | 0.011977357 | 0.771811571 |
| dbCAN_2:Hotpep-dbCAN_2:DIAMOND | -0.0529375 | -0.081727357 | -0.024147643 | 1.12E-07 |
| dbCAN_3-dbCAN_2:DIAMOND | 0.028125 | -0.000664857 | 0.056914857 | 0.063069725 |
| dbCAN_3:DIAMOND-dbCAN_2:DIAMOND | 0.022625 | -0.006164857 | 0.051414857 | 0.305684702 |
| dbCAN_3:eCAMI-dbCAN_2:DIAMOND | -0.031625 | -0.060414857 | -0.002835143 | 0.016971906 |
| dbCAN_3:HMMER-dbCAN_2:DIAMOND | -0.0116875 | -0.040477357 | 0.017102357 | 0.981103985 |
| dbCAN_4-dbCAN_2:DIAMOND | 0.028125 | -0.000664857 | 0.056914857 | 0.063069725 |
| dbCAN_4:dbCAN-sub-dbCAN_2:DIAMOND | 0.0185625 | -0.010227357 | 0.047352357 | 0.634649599 |
| dbCAN_4:DIAMOND-dbCAN_2:DIAMOND | 0.023 | -0.005789857 | 0.051789857 | 0.280249534 |
| dbCAN_4:HMMER-dbCAN_2:DIAMOND | -0.0106875 | -0.039477357 | 0.018102357 | 0.991191366 |
| dbCAN_2:Hotpep-dbCAN_2:HMMER | -0.036125 | -0.064914857 | -0.007335143 | 0.002318383 |
| dbCAN_3-dbCAN_2:HMMER | 0.0449375 | 0.016147643 | 0.073727357 | 2.00E-05 |
| dbCAN_3:DIAMOND-dbCAN_2:HMMER | 0.0394375 | 0.010647643 | 0.068227357 | 0.00044026 |
| dbCAN_3:eCAMI-dbCAN_2:HMMER | -0.0148125 | -0.043602357 | 0.013977357 | 0.891225925 |
| dbCAN_3:HMMER-dbCAN_2:HMMER | 0.005125 | -0.023664857 | 0.033914857 | 0.999995304 |
| dbCAN_4-dbCAN_2:HMMER | 0.0449375 | 0.016147643 | 0.073727357 | 2.00E-05 |
| dbCAN_4:dbCAN-sub-dbCAN_2:HMMER | 0.035375 | 0.006585143 | 0.064164857 | 0.003302463 |
| dbCAN_4:DIAMOND-dbCAN_2:HMMER | 0.0398125 | 0.011022643 | 0.068602357 | 0.00036122 |
| dbCAN_4:HMMER-dbCAN_2:HMMER | 0.006125 | -0.022664857 | 0.034914857 | 0.999966355 |
| dbCAN_3-dbCAN_2:Hotpep | 0.0810625 | 0.052272643 | 0.109852357 | 4.10E-13 |
| dbCAN_3:DIAMOND-dbCAN_2:Hotpep | 0.0755625 | 0.046772643 | 0.104352357 | 3.96E-13 |
| dbCAN_3:eCAMI-dbCAN_2:Hotpep | 0.0213125 | -0.007477357 | 0.050102357 | 0.403866958 |
| dbCAN_3:HMMER-dbCAN_2:Hotpep | 0.04125 | 0.012460143 | 0.070039857 | 0.000166182 |
| dbCAN_4-dbCAN_2:Hotpep | 0.0810625 | 0.052272643 | 0.109852357 | 4.10E-13 |
| dbCAN_4:dbCAN-sub-dbCAN_2:Hotpep | 0.0715 | 0.042710143 | 0.100289857 | 4.71E-13 |
| dbCAN_4:DIAMOND-dbCAN_2:Hotpep | 0.0759375 | 0.047147643 | 0.104727357 | 3.98E-13 |
| dbCAN_4:HMMER-dbCAN_2:Hotpep | 0.04225 | 0.013460143 | 0.071039857 | 9.53E-05 |
| dbCAN_3:DIAMOND-dbCAN_3 | -0.0055 | -0.034289857 | 0.023289857 | 0.999989683 |
| dbCAN_3:eCAMI-dbCAN_3 | -0.05975 | -0.088539857 | -0.030960143 | 7.50E-10 |
| dbCAN_3:HMMER-dbCAN_3 | -0.0398125 | -0.068602357 | -0.011022643 | 0.00036122 |
| dbCAN_4-dbCAN_3 | 0 | -0.028789857 | 0.028789857 | 1 |
| dbCAN_4:dbCAN-sub-dbCAN_3 | -0.0095625 | -0.038352357 | 0.019227357 | 0.996799164 |
| dbCAN_4:DIAMOND-dbCAN_3 | -0.005125 | -0.033914857 | 0.023664857 | 0.999995304 |
| dbCAN_4:HMMER-dbCAN_3 | -0.0388125 | -0.067602357 | -0.010022643 | 0.000609605 |
| dbCAN_3:eCAMI-dbCAN_3:DIAMOND | -0.05425 | -0.083039857 | -0.025460143 | 4.47E-08 |
| dbCAN_3:HMMER-dbCAN_3:DIAMOND | -0.0343125 | -0.063102357 | -0.005522643 | 0.00537186 |
| dbCAN_4-dbCAN_3:DIAMOND | 0.0055 | -0.023289857 | 0.034289857 | 0.999989683 |
| dbCAN_4:dbCAN-sub-dbCAN_3:DIAMOND | -0.0040625 | -0.032852357 | 0.024727357 | 0.999999665 |
| dbCAN_4:DIAMOND-dbCAN_3:DIAMOND | 0.000375 | -0.028414857 | 0.029164857 | 1 |
| dbCAN_4:HMMER-dbCAN_3:DIAMOND | -0.0333125 | -0.062102357 | -0.004522643 | 0.008355444 |
| dbCAN_3:HMMER-dbCAN_3:eCAMI | 0.0199375 | -0.008852357 | 0.048727357 | 0.517748566 |
| dbCAN_4-dbCAN_3:eCAMI | 0.05975 | 0.030960143 | 0.088539857 | 7.50E-10 |
| dbCAN_4:dbCAN-sub-dbCAN_3:eCAMI | 0.0501875 | 0.021397643 | 0.078977357 | 7.29E-07 |
| dbCAN_4:DIAMOND-dbCAN_3:eCAMI | 0.054625 | 0.025835143 | 0.083414857 | 3.42E-08 |
| dbCAN_4:HMMER-dbCAN_3:eCAMI | 0.0209375 | -0.007852357 | 0.049727357 | 0.434097668 |
| dbCAN_4-dbCAN_3:HMMER | 0.0398125 | 0.011022643 | 0.068602357 | 0.00036122 |
| dbCAN_4:dbCAN-sub-dbCAN_3:HMMER | 0.03025 | 0.001460143 | 0.059039857 | 0.029177093 |
| dbCAN_4:DIAMOND-dbCAN_3:HMMER | 0.0346875 | 0.005897643 | 0.063477357 | 0.004533344 |
| dbCAN_4:HMMER-dbCAN_3:HMMER | 0.001 | -0.027789857 | 0.029789857 | 1 |
| dbCAN_4:dbCAN-sub-dbCAN_4 | -0.0095625 | -0.038352357 | 0.019227357 | 0.996799164 |
| dbCAN_4:DIAMOND-dbCAN_4 | -0.005125 | -0.033914857 | 0.023664857 | 0.999995304 |
| dbCAN_4:HMMER-dbCAN_4 | -0.0388125 | -0.067602357 | -0.010022643 | 0.000609605 |
| dbCAN_4:DIAMOND-dbCAN_4:dbCAN-sub | 0.0044375 | -0.024352357 | 0.033227357 | 0.999999079 |
| dbCAN_4:HMMER-dbCAN_4:dbCAN-sub | -0.02925 | -0.058039857 | -0.000460143 | 0.042371328 |
| dbCAN_4:HMMER-dbCAN_4:DIAMOND | -0.0336875 | -0.062477357 | -0.004897643 | 0.007093116 |

# 3 Taxonomic performance of CAZyme/non-CAZyme classification

## 3.1 Summary of taxonomic kingdom performance of CAZyme/non-CAZyme classification

### SI Table 7: Summary table of binary CAZyme/non-CAZyme classification of bacterial and eukaryotic protein sequences (overleaf)

SI table 10 lists the statistical parameter values for the binary CAZyme/non-CAZyme classification of protein sequences from bacteria and eukaryotes, reporting the mean, standard deviation,and lower and upper 95% confidence intervals (CIs) across all test sets from bacteria, eukaryotes, and both kingdoms, for each statistical parameter.

| Statistical parameter | Prediction Tool | Kingdom | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | Bacteria | | | | All | | | | Eukaryote | | | |
| | | Lower CI | Mean | SD | Upper CI | Lower CI | Mean | SD | Upper CI | Lower CI | Mean | SD | Upper CI |
| **Specificity** | CUPP | 0.980 | 0.989 | 0.028 | 0.998 | 0.971 | 0.982 | 0.049 | 0.993 | 0.955 | 0.975 | 0.063 | 0.995 |
| | dbCAN_2 | 0.972 | 0.983 | 0.035 | 0.994 | 0.966 | 0.977 | 0.053 | 0.989 | 0.951 | 0.972 | 0.066 | 0.993 |
| | dbCAN_2:DIAMOND | 0.968 | 0.979 | 0.035 | 0.990 | 0.962 | 0.974 | 0.054 | 0.986 | 0.948 | 0.970 | 0.068 | 0.992 |
| | dbCAN_2:HMMER | 0.977 | 0.986 | 0.029 | 0.995 | 0.967 | 0.979 | 0.052 | 0.990 | 0.950 | 0.972 | 0.067 | 0.993 |
| | dbCAN_2:Hotpep | 0.968 | 0.979 | 0.034 | 0.990 | 0.966 | 0.976 | 0.046 | 0.986 | 0.955 | 0.973 | 0.055 | 0.991 |
| | dbCAN_3 | 0.956 | 0.969 | 0.040 | 0.981 | 0.942 | 0.959 | 0.077 | 0.976 | 0.918 | 0.950 | 0.101 | 0.982 |
| | dbCAN_3:DIAMOND | 0.967 | 0.979 | 0.036 | 0.990 | 0.950 | 0.967 | 0.074 | 0.983 | 0.924 | 0.955 | 0.097 | 0.985 |
| | dbCAN_3:HMMER | 0.978 | 0.987 | 0.028 | 0.996 | 0.967 | 0.979 | 0.052 | 0.990 | 0.949 | 0.971 | 0.067 | 0.992 |
| | dbCAN_3:eCAMI | 0.966 | 0.977 | 0.035 | 0.988 | 0.965 | 0.975 | 0.048 | 0.986 | 0.955 | 0.973 | 0.058 | 0.992 |
| | dbCAN_4 | 0.962 | 0.974 | 0.038 | 0.986 | 0.945 | 0.962 | 0.076 | 0.979 | 0.918 | 0.951 | 0.100 | 0.983 |
| | dbCAN_4:DIAMOND | 0.968 | 0.980 | 0.035 | 0.991 | 0.950 | 0.967 | 0.074 | 0.983 | 0.922 | 0.954 | 0.098 | 0.985 |
| | dbCAN_4:HMMER | 0.978 | 0.987 | 0.028 | 0.996 | 0.967 | 0.979 | 0.052 | 0.990 | 0.949 | 0.971 | 0.067 | 0.992 |
| | dbCAN_4:dbCAN-sub | 0.978 | 0.987 | 0.027 | 0.995 | 0.966 | 0.978 | 0.053 | 0.989 | 0.946 | 0.968 | 0.069 | 0.990 |
| **Sensitivity** | CUPP | 0.837 | 0.862 | 0.079 | 0.887 | 0.837 | 0.854 | 0.074 | 0.870 | 0.824 | 0.846 | 0.070 | 0.868 |
| | dbCAN_2 | 0.869 | 0.905 | 0.114 | 0.942 | 0.877 | 0.901 | 0.108 | 0.925 | 0.864 | 0.897 | 0.103 | 0.930 |
| | dbCAN_2:DIAMOND | 0.875 | 0.918 | 0.135 | 0.961 | 0.887 | 0.916 | 0.129 | 0.944 | 0.874 | 0.913 | 0.124 | 0.953 |
| | dbCAN_2:HMMER | 0.835 | 0.870 | 0.108 | 0.904 | 0.860 | 0.878 | 0.081 | 0.896 | 0.874 | 0.886 | 0.038 | 0.898 |
| | dbCAN_2:Hotpep | 0.782 | 0.820 | 0.116 | 0.857 | 0.780 | 0.808 | 0.127 | 0.837 | 0.753 | 0.797 | 0.137 | 0.841 |
| | dbCAN_3 | 0.964 | 0.983 | 0.060 | 1.003 | 0.977 | 0.987 | 0.044 | 0.997 | 0.987 | 0.991 | 0.013 | 0.995 |
| | dbCAN_3:DIAMOND | 0.921 | 0.956 | 0.109 | 0.991 | 0.951 | 0.969 | 0.082 | 0.987 | 0.969 | 0.982 | 0.039 | 0.994 |
| | dbCAN_3:HMMER | 0.838 | 0.873 | 0.109 | 0.908 | 0.870 | 0.888 | 0.082 | 0.906 | 0.892 | 0.903 | 0.035 | 0.915 |
| | dbCAN_3:eCAMI | 0.843 | 0.883 | 0.126 | 0.923 | 0.823 | 0.852 | 0.128 | 0.880 | 0.781 | 0.820 | 0.123 | 0.859 |
| | dbCAN_4 | 0.946 | 0.975 | 0.090 | 1.003 | 0.970 | 0.984 | 0.064 | 0.998 | 0.991 | 0.994 | 0.008 | 0.996 |
| | dbCAN_4:DIAMOND | 0.913 | 0.950 | 0.115 | 0.987 | 0.951 | 0.969 | 0.084 | 0.988 | 0.985 | 0.989 | 0.011 | 0.993 |
| | dbCAN_4:HMMER | 0.841 | 0.877 | 0.110 | 0.912 | 0.872 | 0.890 | 0.082 | 0.908 | 0.893 | 0.904 | 0.034 | 0.914 |
| | dbCAN_4:dbCAN-sub | 0.920 | 0.952 | 0.101 | 0.984 | 0.934 | 0.950 | 0.073 | 0.966 | 0.940 | 0.948 | 0.023 | 0.955 |
| **Precision** | CUPP | 0.979 | 0.988 | 0.028 | 0.997 | 0.971 | 0.981 | 0.045 | 0.991 | 0.956 | 0.975 | 0.057 | 0.993 |
| | dbCAN_2 | 0.971 | 0.982 | 0.035 | 0.994 | 0.968 | 0.978 | 0.045 | 0.988 | 0.956 | 0.973 | 0.054 | 0.991 |
| | dbCAN_2:DIAMOND | 0.968 | 0.979 | 0.035 | 0.990 | 0.966 | 0.976 | 0.045 | 0.986 | 0.956 | 0.973 | 0.053 | 0.990 |
| | dbCAN_2:HMMER | 0.975 | 0.985 | 0.030 | 0.994 | 0.969 | 0.979 | 0.045 | 0.989 | 0.955 | 0.973 | 0.055 | 0.991 |
| | dbCAN_2:Hotpep | 0.964 | 0.976 | 0.038 | 0.988 | 0.960 | 0.972 | 0.053 | 0.984 | 0.947 | 0.968 | 0.066 | 0.989 |
| | dbCAN_3 | 0.959 | 0.970 | 0.036 | 0.982 | 0.952 | 0.964 | 0.056 | 0.977 | 0.936 | 0.959 | 0.070 | 0.981 |
| | dbCAN_3:DIAMOND | 0.969 | 0.979 | 0.033 | 0.990 | 0.959 | 0.971 | 0.054 | 0.983 | 0.940 | 0.962 | 0.069 | 0.984 |
| | dbCAN_3:HMMER | 0.976 | 0.986 | 0.030 | 0.995 | 0.969 | 0.979 | 0.045 | 0.989 | 0.955 | 0.973 | 0.056 | 0.990 |
| | dbCAN_3:eCAMI | 0.965 | 0.976 | 0.035 | 0.988 | 0.964 | 0.974 | 0.046 | 0.984 | 0.953 | 0.971 | 0.056 | 0.989 |
| | dbCAN_4 | 0.964 | 0.976 | 0.035 | 0.987 | 0.955 | 0.967 | 0.055 | 0.980 | 0.937 | 0.959 | 0.070 | 0.981 |
| | dbCAN_4:DIAMOND | 0.970 | 0.981 | 0.033 | 0.991 | 0.959 | 0.971 | 0.055 | 0.983 | 0.939 | 0.962 | 0.069 | 0.984 |
| | dbCAN_4:HMMER | 0.976 | 0.986 | 0.030 | 0.995 | 0.969 | 0.979 | 0.045 | 0.989 | 0.955 | 0.973 | 0.056 | 0.990 |
| | dbCAN_4:dbCAN-sub | 0.979 | 0.987 | 0.026 | 0.995 | 0.970 | 0.979 | 0.043 | 0.989 | 0.954 | 0.971 | 0.055 | 0.989 |
| **F1-score** | CUPP | 0.902 | 0.918 | 0.052 | 0.935 | 0.899 | 0.911 | 0.052 | 0.923 | 0.887 | 0.904 | 0.052 | 0.920 |
| | dbCAN_2 | 0.911 | 0.937 | 0.084 | 0.964 | 0.916 | 0.933 | 0.079 | 0.951 | 0.905 | 0.929 | 0.075 | 0.953 |
| | dbCAN_2:DIAMOND | 0.909 | 0.941 | 0.099 | 0.972 | 0.917 | 0.938 | 0.092 | 0.959 | 0.908 | 0.935 | 0.087 | 0.963 |
| | dbCAN_2:HMMER | 0.892 | 0.919 | 0.084 | 0.946 | 0.908 | 0.923 | 0.064 | 0.937 | 0.915 | 0.926 | 0.035 | 0.937 |
| | dbCAN_2:Hotpep | 0.862 | 0.886 | 0.076 | 0.910 | 0.857 | 0.877 | 0.092 | 0.897 | 0.834 | 0.868 | 0.105 | 0.902 |
| | dbCAN_3 | 0.962 | 0.975 | 0.041 | 0.988 | 0.965 | 0.974 | 0.041 | 0.983 | 0.960 | 0.973 | 0.042 | 0.986 |
| | dbCAN_3:DIAMOND | 0.937 | 0.963 | 0.082 | 0.989 | 0.952 | 0.966 | 0.066 | 0.981 | 0.955 | 0.970 | 0.046 | 0.985 |
| | dbCAN_3:HMMER | 0.894 | 0.921 | 0.085 | 0.948 | 0.914 | 0.928 | 0.065 | 0.943 | 0.924 | 0.935 | 0.035 | 0.947 |
| | dbCAN_3:eCAMI | 0.896 | 0.922 | 0.080 | 0.948 | 0.883 | 0.903 | 0.087 | 0.922 | 0.854 | 0.883 | 0.090 | 0.912 |
| | dbCAN_4 | 0.952 | 0.972 | 0.063 | 0.992 | 0.961 | 0.973 | 0.053 | 0.985 | 0.961 | 0.975 | 0.042 | 0.988 |
| | dbCAN_4:DIAMOND | 0.931 | 0.959 | 0.089 | 0.988 | 0.951 | 0.967 | 0.069 | 0.982 | 0.960 | 0.974 | 0.042 | 0.987 |
| | dbCAN_4:HMMER | 0.896 | 0.923 | 0.085 | 0.950 | 0.915 | 0.929 | 0.065 | 0.944 | 0.925 | 0.936 | 0.034 | 0.947 |
| | dbCAN_4:dbCAN-sub | 0.942 | 0.965 | 0.074 | 0.989 | 0.949 | 0.962 | 0.056 | 0.974 | 0.948 | 0.958 | 0.031 | 0.968 |
| **Accuracy** | CUPP | 0.911 | 0.925 | 0.043 | 0.939 | 0.908 | 0.918 | 0.046 | 0.928 | 0.895 | 0.910 | 0.048 | 0.926 |
| | dbCAN_2 | 0.925 | 0.944 | 0.060 | 0.963 | 0.926 | 0.939 | 0.060 | 0.953 | 0.915 | 0.935 | 0.059 | 0.954 |
| | dbCAN_2:DIAMOND | 0.927 | 0.949 | 0.069 | 0.971 | 0.930 | 0.945 | 0.067 | 0.960 | 0.920 | 0.942 | 0.067 | 0.963 |
| | dbCAN_2:HMMER | 0.910 | 0.928 | 0.056 | 0.946 | 0.918 | 0.928 | 0.047 | 0.939 | 0.917 | 0.929 | 0.037 | 0.941 |
| | dbCAN_2:Hotpep | 0.880 | 0.899 | 0.060 | 0.918 | 0.876 | 0.892 | 0.071 | 0.908 | 0.859 | 0.885 | 0.081 | 0.911 |
| | dbCAN_3 | 0.965 | 0.976 | 0.035 | 0.987 | 0.963 | 0.973 | 0.044 | 0.983 | 0.954 | 0.971 | 0.051 | 0.987 |
| | dbCAN_3:DIAMOND | 0.949 | 0.967 | 0.056 | 0.985 | 0.956 | 0.968 | 0.054 | 0.980 | 0.951 | 0.968 | 0.053 | 0.985 |
| | dbCAN_3:HMMER | 0.912 | 0.930 | 0.057 | 0.948 | 0.923 | 0.933 | 0.048 | 0.944 | 0.925 | 0.937 | 0.038 | 0.949 |
| | dbCAN_3:eCAMI | 0.910 | 0.930 | 0.064 | 0.951 | 0.898 | 0.913 | 0.068 | 0.929 | 0.874 | 0.897 | 0.069 | 0.919 |
| | dbCAN_4 | 0.959 | 0.974 | 0.047 | 0.990 | 0.962 | 0.973 | 0.049 | 0.984 | 0.956 | 0.972 | 0.051 | 0.988 |
| | dbCAN_4:DIAMOND | 0.946 | 0.965 | 0.059 | 0.984 | 0.956 | 0.968 | 0.055 | 0.980 | 0.955 | 0.971 | 0.050 | 0.988 |
| | dbCAN_4:HMMER | 0.913 | 0.932 | 0.057 | 0.950 | 0.924 | 0.934 | 0.048 | 0.945 | 0.925 | 0.937 | 0.038 | 0.949 |
| | dbCAN_4:dbCAN-sub | 0.952 | 0.969 | 0.053 | 0.986 | 0.954 | 0.964 | 0.045 | 0.974 | 0.947 | 0.958 | 0.035 | 0.969 |

**SI Figure 3: 95% confidence intervals of the specificity of CAZyme/non-CAZyme classification per taxonomic kingdom**

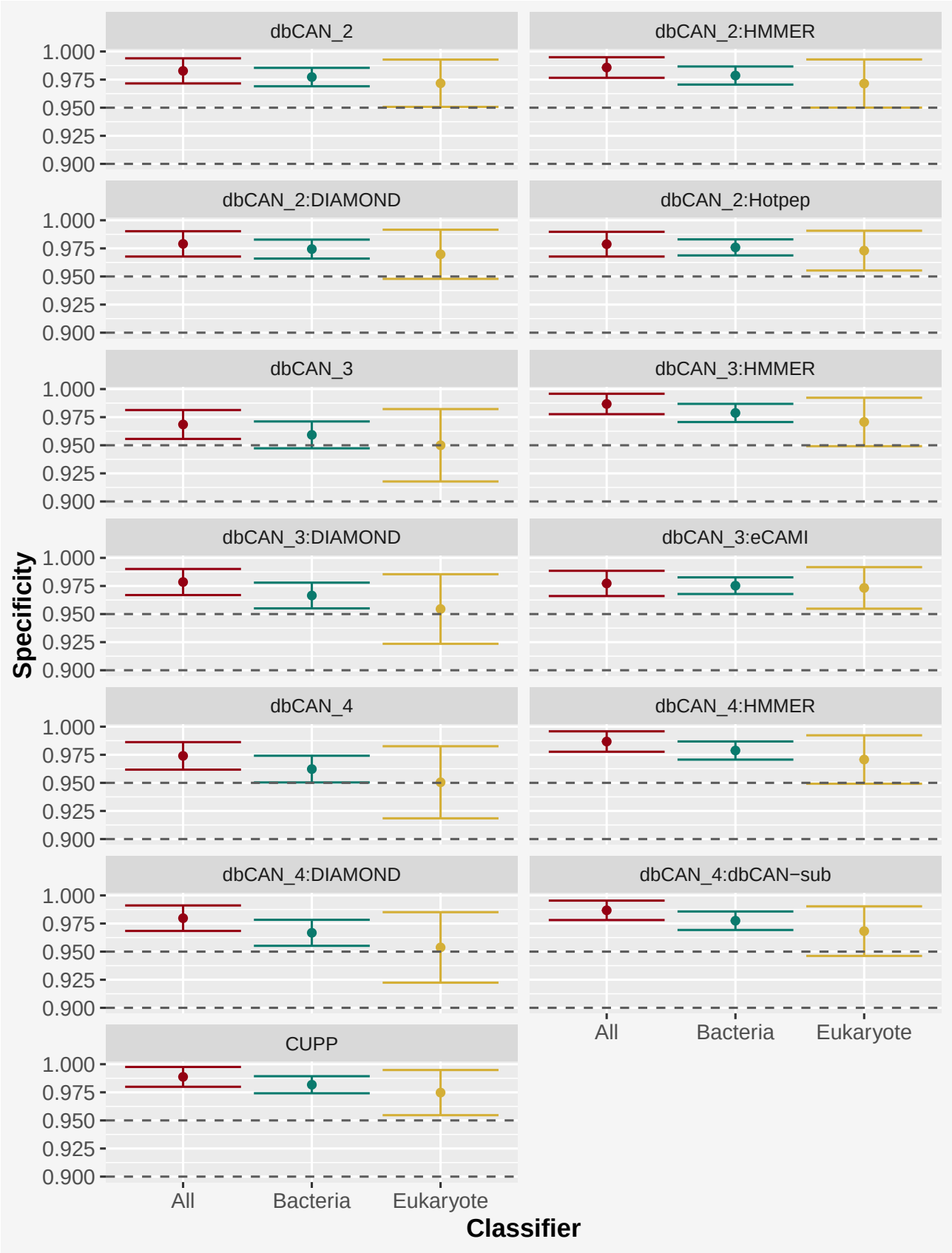



Figure 3: Mean and 95% confidence interval of the specificity across all test sets per taxonomic kingdom (Bacteria shaded green; Eukaryote shaded Yellow; Both/All shaded red) for the CAZyme/non-CAZyme classification of protein sequences.

**SI Figure 4: The specificity of CAZyme/non-CAZyme classification per taxonomic kingdom for each test set**

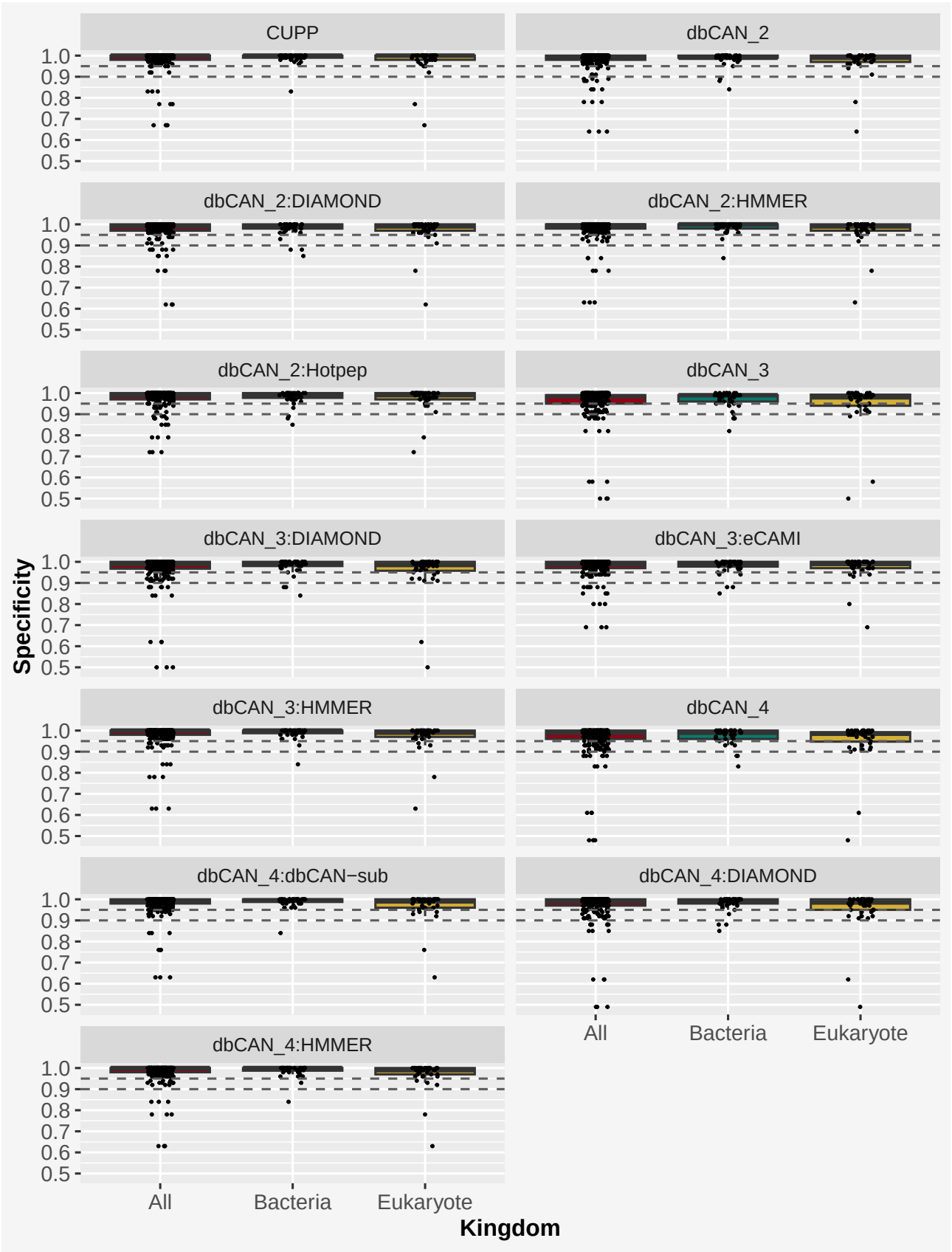



Figure 4: One-dimensional scatter plot overlaying a box plot, where each point represents the specificity of unique test set, for the binary CAZyme/non-CAZyme classification of protein sequences, per taxonomic kingdom.

**SI Figure 5: 95% confidence intervals of the sensitivity of CAZyme/non-CAZyme classification per taxonomic kingdom**

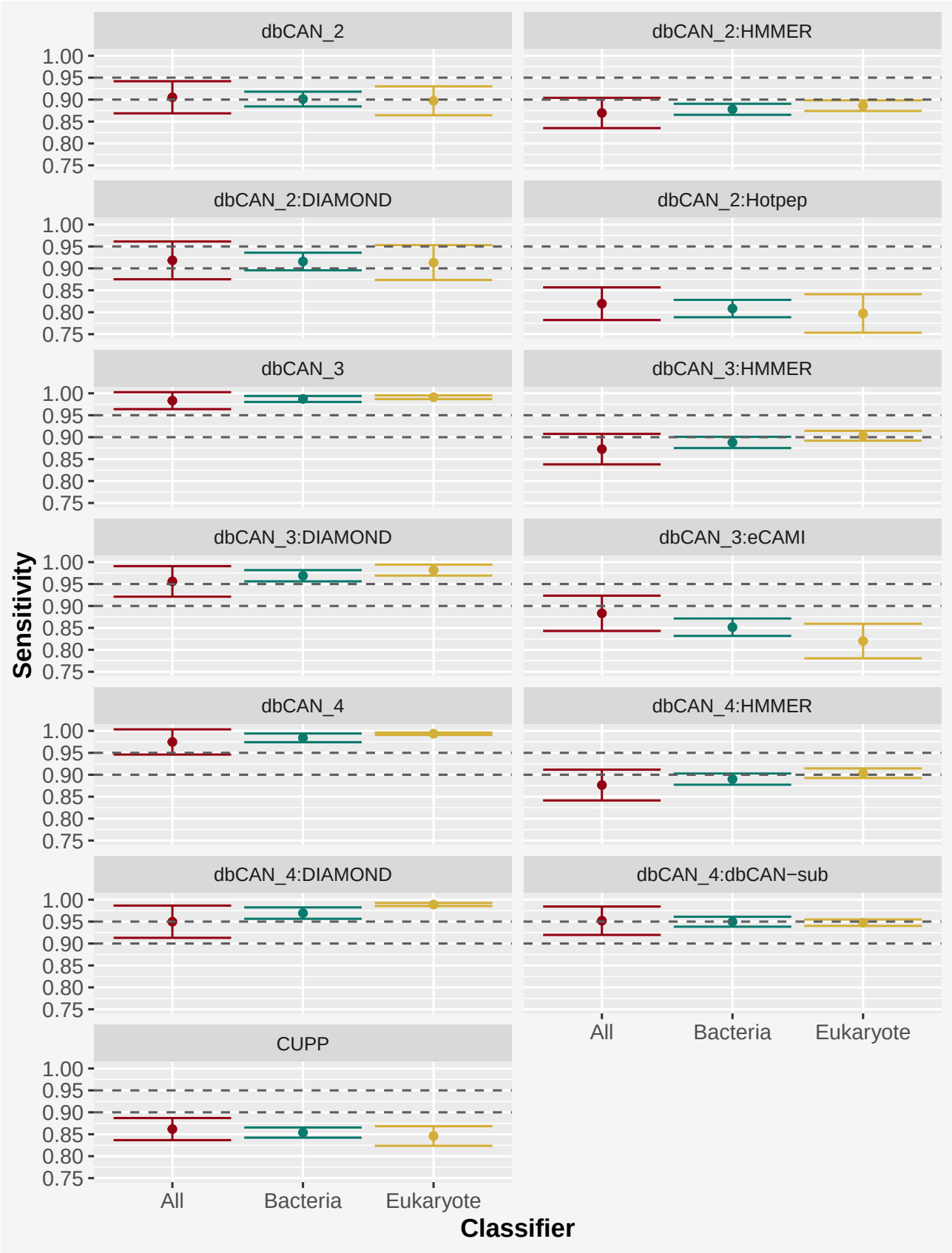



Figure 5: Mean and 95% confidence interval of the sensitivity across all test sets per taxonomic kingdom (Bacteria shaded green; Eukaryote shaded Yellow; Both/All shaded red) for the CAZyme/non-CAZyme classification of protein sequences.

**SI Figure 6: The sensitivity of CAZyme/non-CAZyme classification per taxonomic kingdom for each test set**

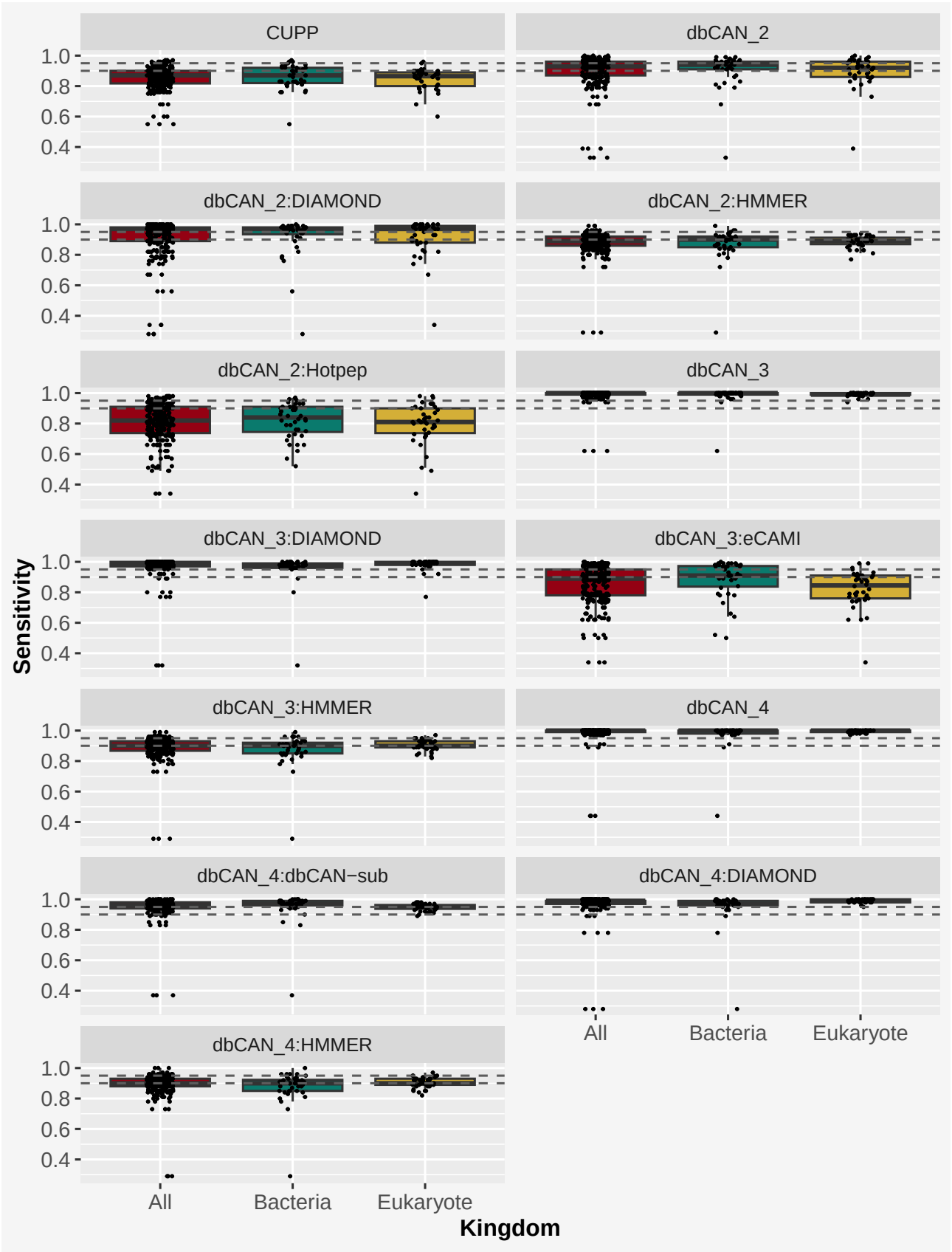



Figure 6: One dimensional scatter plot overlaying a box plot, where each point represents the sensitivity of unique test set, for the binary CAZyme/non-CAZyme classification of protein sequences, per taxonomic kingdom.

**SI Figure 7: 95% confidence intervals of the precision of CAZyme/non-CAZyme classification per taxonomic kingdom**

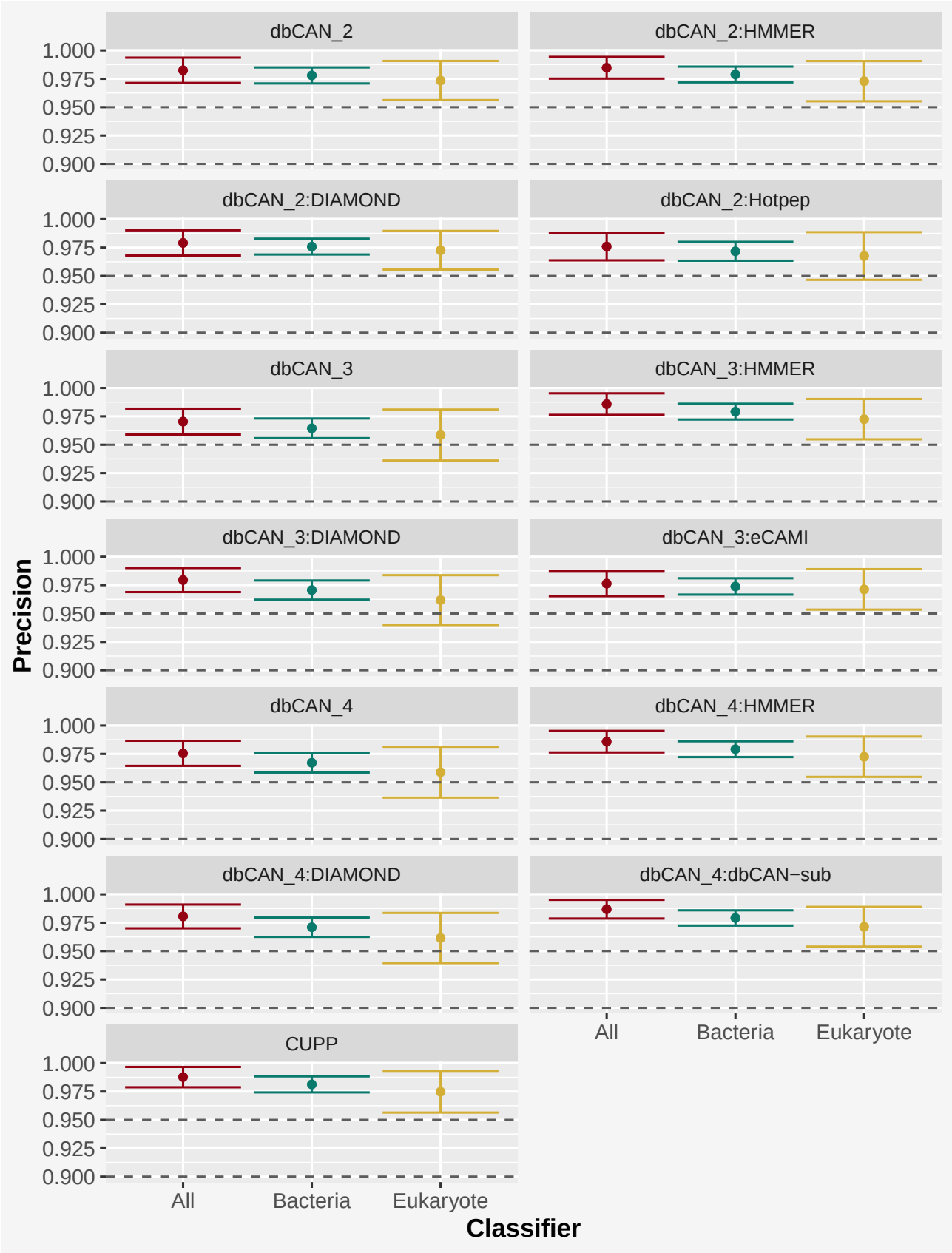



Figure 7: Mean and 95% confidence interval of the precision across all test sets per taxonomic kingdom (Bacteria shaded green; Eukaryote shaded Yellow; Both/All shaded red) for the CAZyme/non-CAZyme classification of protein sequences.

**SI Figure 8: The precision of CAZyme/non-CAZyme classification per taxonomic kingdom for each test set**

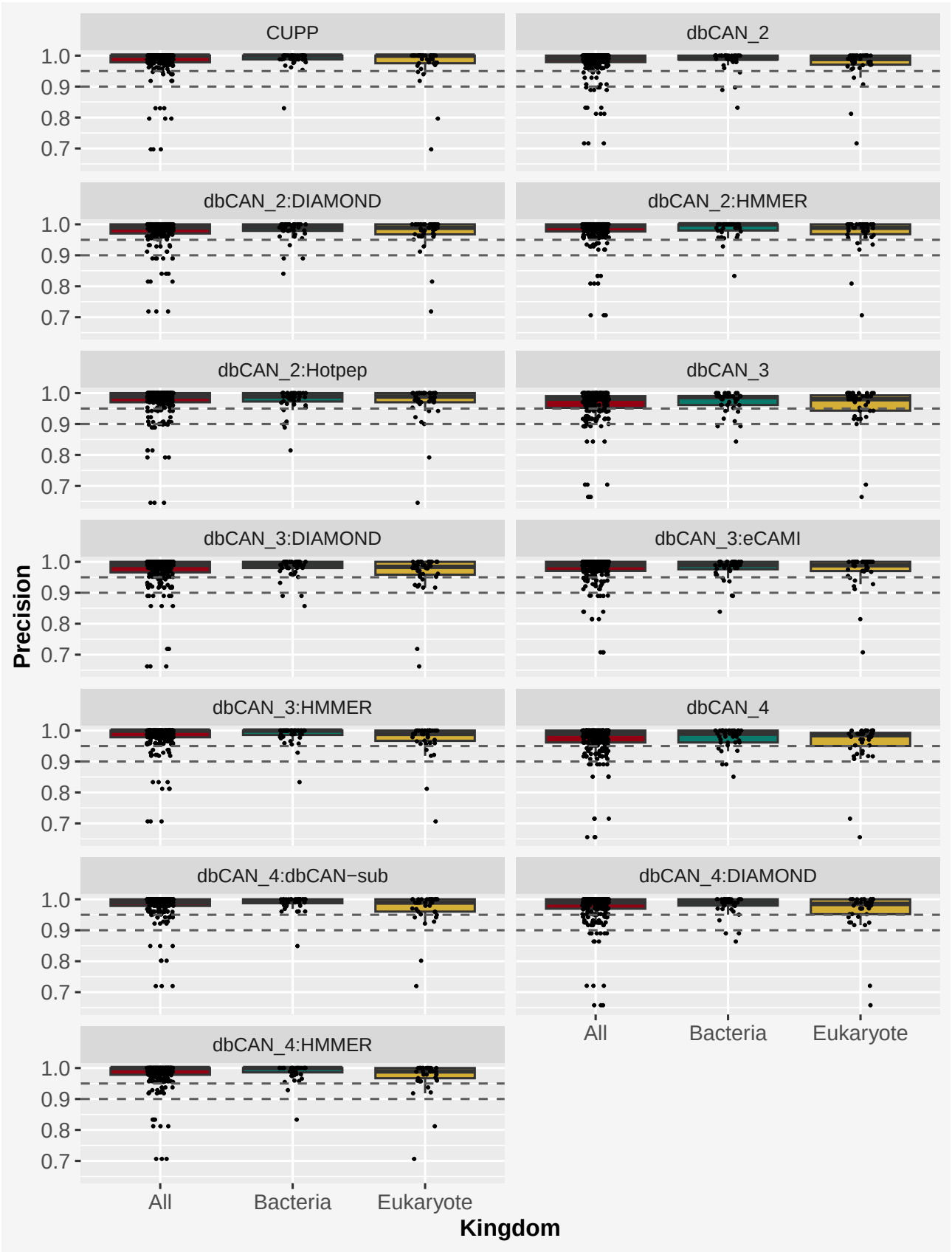



Figure 8: One dimensional scatter plot overlaying a box plot, where each point represents the precision of unique test set, for the binary CAZyme/non-CAZyme classification of protein sequences, per taxonomic kingdom.

**SI Figure 9: 95% confidence intervals of the F1-score of CAZyme/non-CAZyme classification per taxonomic kingdom**

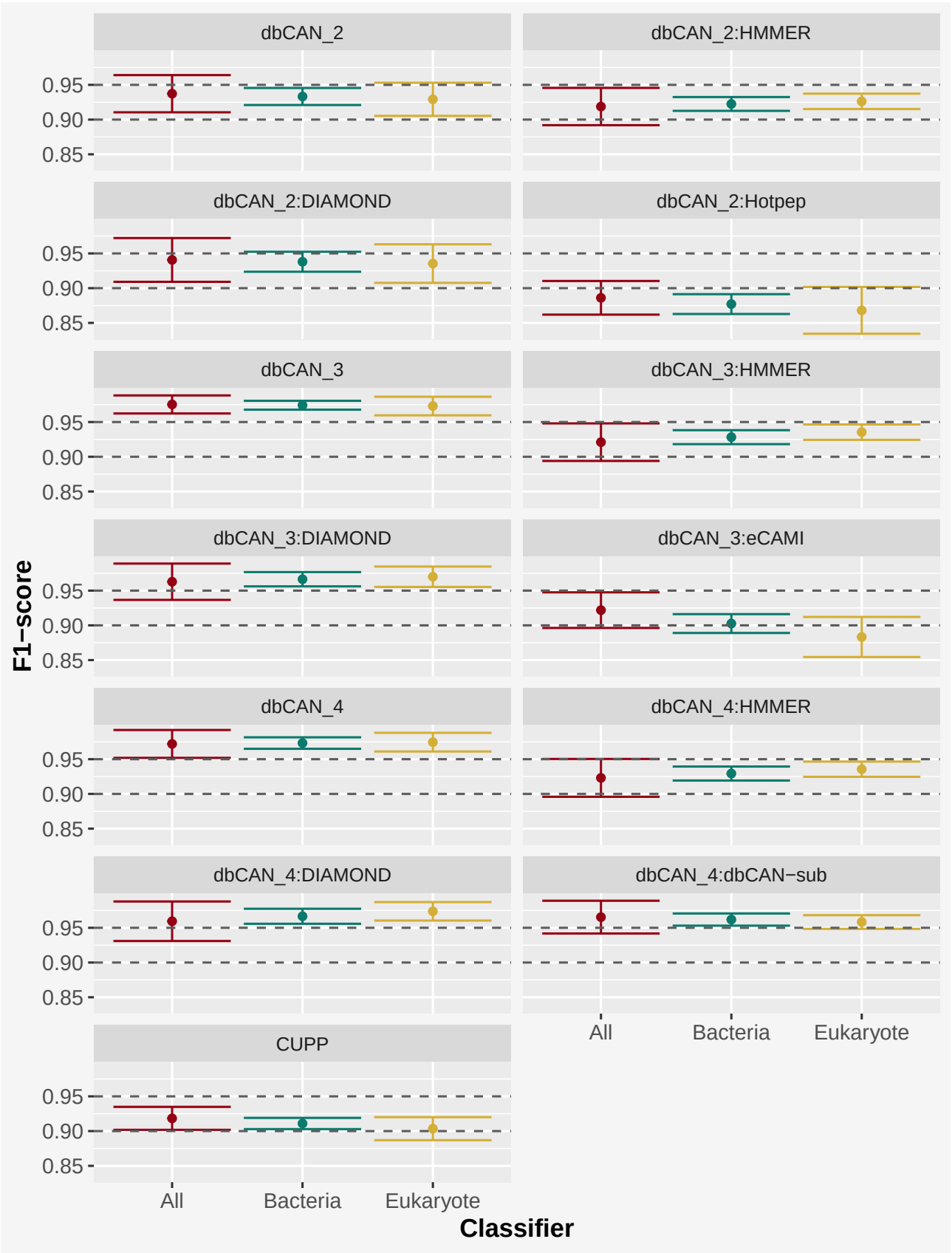



Figure 9: Mean and 95% confidence interval of the F1-score across all test sets per taxonomic kingdom (Bacteria shaded green; Eukaryote shaded Yellow; Both/All shaded red) for the CAZyme/non-CAZyme classification of protein sequences.

**SI Figure 10: The F1-score of CAZyme/non-CAZyme classification per taxonomic kingdom for each test set**

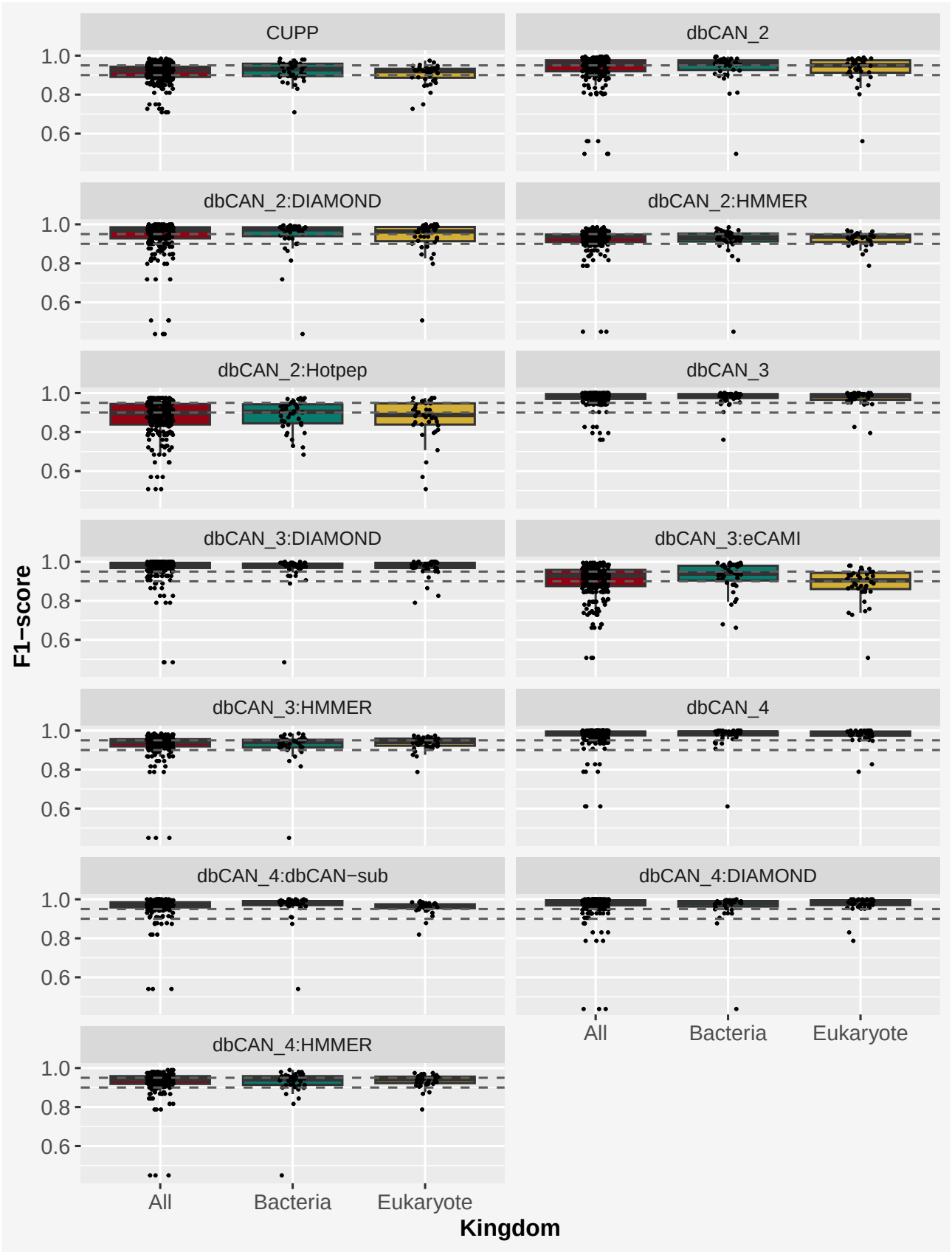



Figure 10: One dimensional scatter plot overlaying a box plot, where each point represents the F1-score of unique test set, for the binary CAZyme/non-CAZyme classification of protein sequences, per taxonomic kingdom.

**SI Figure 11: 95% confidence intervals of the accuracy of CAZyme/non-CAZyme classification per taxonomic kingdom**

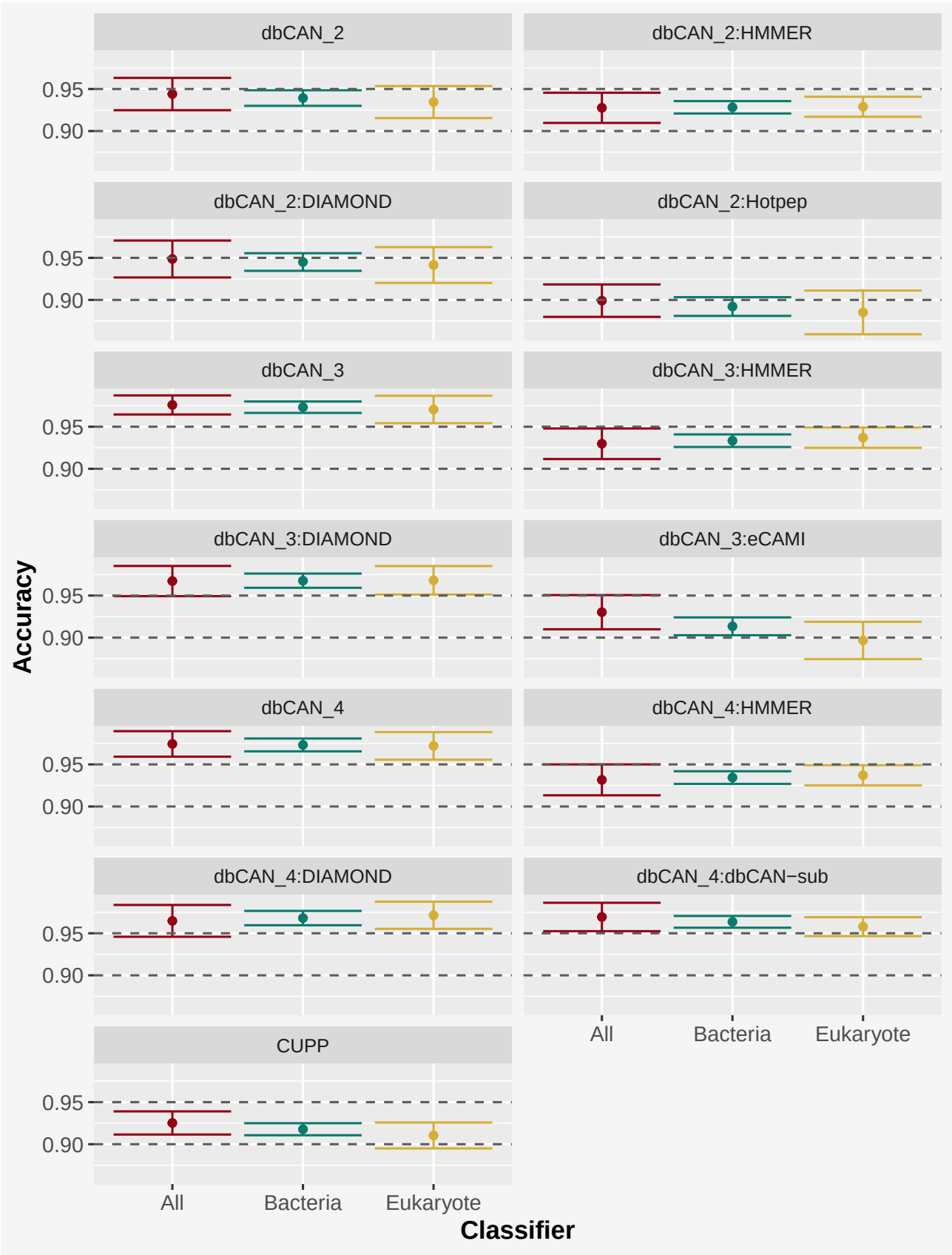



Figure 11: Mean and 95% confidence interval of the accuracy across all test sets per taxonomic kingdom (Bacteria shaded green; Eukaryote shaded Yellow; Both/All shaded red) for the CAZyme/non-CAZyme classification of protein sequences.

**SI Figure 12: The accuracy of CAZyme/non-CAZyme classification per taxonomic kingdom for each test set**

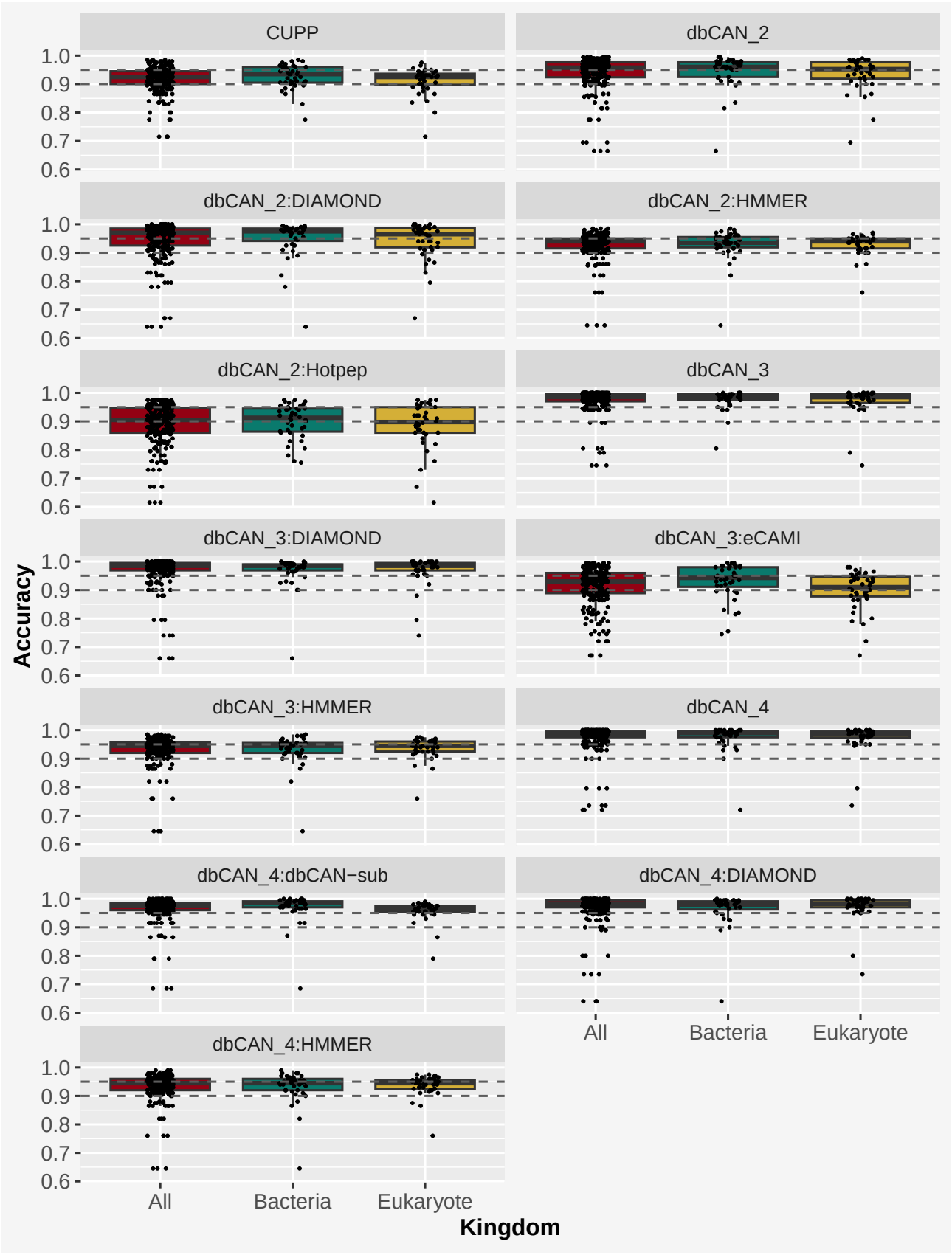



Figure 12: One dimensional scatter plot overlaying a box plot, where each point represents the accuracy of unique test set, for the binary CAZyme/non-CAZyme classification of protein sequences, per taxonomic kingdom.

## 3.2 Output of testing for statistically significant difference in performance between the taxonomic kingdoms

### SI table 8: Output of a two-way ANOVA testing for statistically difference between prediction tools and taxonomic kingdoms for specificity of CAZyme/non-CAZyme classification

| Term | DF | SumSq | MeanSq | Statistic | P-value |
|---|---|---|---|---|---|
| Tax_group | 2 | 0.062001 | 0.031 | 8.795037 | 0.000157 |
| Prediction_tool | 12 | 0.098023 | 0.008169 | 2.317476 | 0.00614 |
| Tax_group:Prediction_tool | 24 | 0.011108 | 0.000463 | 0.131308 | 1 |
| Residuals | 2041 | 7.194044 | 0.003525 | NA | NA |

### SI table 9: Output of Tukey HSD test following a two-way ANOVA testing for statistically difference between prediction tools and taxonomic kingdoms for specificity of CAZyme/non-CAZyme classification

Reporting the output of a Tukey HSD test following a two-way ANOVA testing for statistically difference between prediction tools and taxonomic kingdoms for specificity of CAZyme/non-CAZyme classification. Specifically reporting the mean difference (MeanDiff), lower and upper 95% confidence interval (CI), and the adjusted p-value (P.Adj). Additionally, only the results comparing the output of the same tool for different taxonomic kingdoms are presented. The comprehensive output with all pairwise comparisons is available in the online repository.

| Kingdom:Classifier 1 - Kingdom:Classifier 2 | MeanDiff | Lower CI | Upper CI | P.Adj |
|---|---|---|---|---|
| Bacteria:CUPP-All:CUPP | 0.007 | -0.03763 | 0.051626 | 1 |
| Eukaryote:CUPP-All:CUPP | -0.007 | -0.05163 | 0.037626 | 1 |
| Eukaryote:CUPP-Bacteria:CUPP | -0.014 | -0.06553 | 0.03753 | 1 |
| Bacteria:dbCAN_2-All:dbCAN_2 | 0.0055 | -0.03913 | 0.050126 | 1 |
| Eukaryote:dbCAN_2-All:dbCAN_2 | -0.0055 | -0.05013 | 0.039126 | 1 |
| Eukaryote:dbCAN_2-Bacteria:dbCAN_2 | -0.011 | -0.06253 | 0.04053 | 1 |
| Bacteria:dbCAN_2:DIAMOND-All:dbCAN_2:DIAMOND | 0.004625 | -0.04 | 0.049251 | 1 |
| Eukaryote:dbCAN_2:DIAMOND-All:dbCAN_2:DIAMOND | -0.00463 | -0.04925 | 0.040001 | 1 |
| Eukaryote:dbCAN_2:DIAMOND-Bacteria:dbCAN_2:DIAMOND | -0.00925 | -0.06078 | 0.04228 | 1 |
| Bacteria:dbCAN_2:HMMER-All:dbCAN_2:HMMER | 0.007125 | -0.0375 | 0.051751 | 1 |
| Eukaryote:dbCAN_2:HMMER-All:dbCAN_2:HMMER | -0.00713 | -0.05175 | 0.037501 | 1 |
| Eukaryote:dbCAN_2:HMMER-Bacteria:dbCAN_2:HMMER | -0.01425 | -0.06578 | 0.03728 | 1 |
| Bacteria:dbCAN_2:Hotpep-All:dbCAN_2:Hotpep | 0.002875 | -0.04175 | 0.047501 | 1 |
| Eukaryote:dbCAN_2:Hotpep-All:dbCAN_2:Hotpep | -0.00288 | -0.0475 | 0.041751 | 1 |
| Eukaryote:dbCAN_2:Hotpep-Bacteria:dbCAN_2:Hotpep | -0.00575 | -0.05728 | 0.04578 | 1 |
| Bacteria:dbCAN_3-All:dbCAN_3 | 0.00925 | -0.03538 | 0.053876 | 1 |
| Eukaryote:dbCAN_3-All:dbCAN_3 | -0.00925 | -0.05388 | 0.035376 | 1 |
| Eukaryote:dbCAN_3-Bacteria:dbCAN_3 | -0.0185 | -0.07003 | 0.03303 | 0.999998 |
| Bacteria:dbCAN_3:DIAMOND-All:dbCAN_3:DIAMOND | 0.012 | -0.03263 | 0.056626 | 1 |
| Eukaryote:dbCAN_3:DIAMOND-All:dbCAN_3:DIAMOND | -0.012 | -0.05663 | 0.032626 | 1 |
| Eukaryote:dbCAN_3:DIAMOND-Bacteria:dbCAN_3:DIAMOND | -0.024 | -0.07553 | 0.02753 | 0.999227 |
| Bacteria:dbCAN_3:eCAMI-All:dbCAN_3:eCAMI | 0.002 | -0.04263 | 0.046626 | 1 |
| Eukaryote:dbCAN_3:eCAMI-All:dbCAN_3:eCAMI | -0.002 | -0.04663 | 0.042626 | 1 |
| Eukaryote:dbCAN_3:eCAMI-Bacteria:dbCAN_3:eCAMI | -0.004 | -0.05553 | 0.04753 | 1 |
| Bacteria:dbCAN_3:HMMER-All:dbCAN_3:HMMER | 0.008 | -0.03663 | 0.052626 | 1 |
| Eukaryote:dbCAN_3:HMMER-All:dbCAN_3:HMMER | -0.008 | -0.05263 | 0.036626 | 1 |
| Eukaryote:dbCAN_3:HMMER-Bacteria:dbCAN_3:HMMER | -0.016 | -0.06753 | 0.03553 | 1 |
| Bacteria:dbCAN_4-All:dbCAN_4 | 0.01175 | -0.03288 | 0.056376 | 1 |
| Eukaryote:dbCAN_4-All:dbCAN_4 | -0.01175 | -0.05638 | 0.032876 | 1 |
| Eukaryote:dbCAN_4-Bacteria:dbCAN_4 | -0.0235 | -0.07503 | 0.02803 | 0.999499 |
| Bacteria:dbCAN_4:dbCAN-sub-All:dbCAN_4:dbCAN-sub | 0.00925 | -0.03538 | 0.053876 | 1 |
| Eukaryote:dbCAN_4:dbCAN-sub-All:dbCAN_4:dbCAN-sub | -0.00925 | -0.05388 | 0.035376 | 1 |
| Eukaryote:dbCAN_4:dbCAN-sub-Bacteria:dbCAN_4:dbCAN-sub | -0.0185 | -0.07003 | 0.03303 | 0.999998 |
| Bacteria:dbCAN_4:DIAMOND-All:dbCAN_4:DIAMOND | 0.013 | -0.03163 | 0.057626 | 1 |
| Eukaryote:dbCAN_4:DIAMOND-All:dbCAN_4:DIAMOND | -0.013 | -0.05763 | 0.031626 | 1 |
| Eukaryote:dbCAN_4:DIAMOND-Bacteria:dbCAN_4:DIAMOND | -0.026 | -0.07753 | 0.02553 | 0.996398 |
| Bacteria:dbCAN_4:HMMER-All:dbCAN_4:HMMER | 0.008 | -0.03663 | 0.052626 | 1 |
| Eukaryote:dbCAN_4:HMMER-All:dbCAN_4:HMMER | -0.008 | -0.05263 | 0.036626 | 1 |
| Eukaryote:dbCAN_4:HMMER-Bacteria:dbCAN_4:HMMER | -0.016 | -0.06753 | 0.03553 | 1 |

## SI table 10: Output of a two-way ANOVA testing for statistically difference between prediction tools and taxonomic kingdoms for precision of CAZyme/non-CAZyme classification

| Term | DF | SumSq | MeanSq | Statistic | P-value |
|---|---|---|---|---|---|
| Tax_group | 2 | 0.039822 | 0.019911 | 8.351809 | 0.000244 |
| Prediction_tool | 12 | 0.053054 | 0.004421 | 1.854503 | 0.035469 |
| Tax_group:Prediction_tool | 24 | 0.004324 | 0.00018 | 0.075565 | 1 |
| Residuals | 2041 | 4.865818 | 0.002384 | NA | NA |

## SI table 11: Output of Tukey HSD test following a two-way ANOVA testing for statistically difference between prediction tools and taxonomic kingdoms for precision of CAZyme/non-CAZyme classification

Reporting the output of a Tukey HSD test following a two-way ANOVA testing for statistically difference between prediction tools and taxonomic kingdoms for precision of CAZyme/non-CAZyme classification. Specifically reporting the mean difference (MeanDiff), lower and upper 95% confidence interval (CI), and the adjusted p-value (P.Adj). Additionally, only the results comparing the output of the same tool for different taxonomic kingdoms are presented. The comprehensive output with all pairwise comparisons is available in the online repository.

| Kingdom:Classifier 1 - Kingdom:Classifier 2 | MeanDiff | Lower CI | Upper CI | P.Adj |
|---|---|---|---|---|
| Bacteria:CUPP-All:CUPP | 0.006443 | -0.03026 | 0.043144 | 1 |
| Eukaryote:CUPP-All:CUPP | -0.00644 | -0.04314 | 0.030259 | 1 |
| Eukaryote:CUPP-Bacteria:CUPP | -0.01289 | -0.05526 | 0.029494 | 1 |
| Bacteria:dbCAN_2-All:dbCAN_2 | 0.004503 | -0.0322 | 0.041204 | 1 |
| Eukaryote:dbCAN_2-All:dbCAN_2 | -0.0045 | -0.0412 | 0.032198 | 1 |
| Eukaryote:dbCAN_2-Bacteria:dbCAN_2 | -0.00901 | -0.05138 | 0.033373 | 1 |
| Bacteria:dbCAN_2:DIAMOND-All:dbCAN_2:DIAMOND | 0.003256 | -0.03344 | 0.039958 | 1 |
| Eukaryote:dbCAN_2:DIAMOND-All:dbCAN_2:DIAMOND | -0.00326 | -0.03996 | 0.033445 | 1 |
| Eukaryote:dbCAN_2:DIAMOND-Bacteria:dbCAN_2:DIAMOND | -0.00651 | -0.04889 | 0.035866 | 1 |
| Bacteria:dbCAN_2:HMMER-All:dbCAN_2:HMMER | 0.005908 | -0.03079 | 0.04261 | 1 |
| Eukaryote:dbCAN_2:HMMER-All:dbCAN_2:HMMER | -0.00591 | -0.04261 | 0.030793 | 1 |
| Eukaryote:dbCAN_2:HMMER-Bacteria:dbCAN_2:HMMER | -0.01182 | -0.0542 | 0.030562 | 1 |
| Bacteria:dbCAN_2:Hotpep-All:dbCAN_2:Hotpep | 0.004174 | -0.03253 | 0.040875 | 1 |
| Eukaryote:dbCAN_2:Hotpep-All:dbCAN_2:Hotpep | -0.00417 | -0.04088 | 0.032527 | 1 |
| Eukaryote:dbCAN_2:Hotpep-Bacteria:dbCAN_2:Hotpep | -0.00835 | -0.05073 | 0.034031 | 1 |
| Bacteria:dbCAN_3-All:dbCAN_3 | 0.005937 | -0.03076 | 0.042639 | 1 |
| Eukaryote:dbCAN_3-All:dbCAN_3 | -0.00594 | -0.04264 | 0.030764 | 1 |
| Eukaryote:dbCAN_3-Bacteria:dbCAN_3 | -0.01187 | -0.05425 | 0.030504 | 1 |
| Bacteria:dbCAN_3:DIAMOND-All:dbCAN_3:DIAMOND | 0.008833 | -0.02787 | 0.045534 | 1 |
| Eukaryote:dbCAN_3:DIAMOND-All:dbCAN_3:DIAMOND | -0.00883 | -0.04553 | 0.027868 | 1 |
| Eukaryote:dbCAN_3:DIAMOND-Bacteria:dbCAN_3:DIAMOND | -0.01767 | -0.06005 | 0.024713 | 0.999931 |
| Bacteria:dbCAN_3:eCAMI-All:dbCAN_3:eCAMI | 0.002574 | -0.03413 | 0.039275 | 1 |
| Eukaryote:dbCAN_3:eCAMI-All:dbCAN_3:eCAMI | -0.00257 | -0.03928 | 0.034128 | 1 |
| Eukaryote:dbCAN_3:eCAMI-Bacteria:dbCAN_3:eCAMI | -0.00515 | -0.04753 | 0.037232 | 1 |
| Bacteria:dbCAN_3:HMMER-All:dbCAN_3:HMMER | 0.006637 | -0.03006 | 0.043339 | 1 |
| Eukaryote:dbCAN_3:HMMER-All:dbCAN_3:HMMER | -0.00664 | -0.04334 | 0.030064 | 1 |
| Eukaryote:dbCAN_3:HMMER-Bacteria:dbCAN_3:HMMER | -0.01327 | -0.05565 | 0.029104 | 1 |
| Bacteria:dbCAN_4-All:dbCAN_4 | 0.008315 | -0.02839 | 0.045016 | 1 |
| Eukaryote:dbCAN_4-All:dbCAN_4 | -0.00831 | -0.04502 | 0.028387 | 1 |
| Eukaryote:dbCAN_4-Bacteria:dbCAN_4 | -0.01663 | -0.05901 | 0.02575 | 0.999984 |
| Bacteria:dbCAN_4:dbCAN-sub-All:dbCAN_4:dbCAN-sub | 0.007715 | -0.02899 | 0.044417 | 1 |
| Eukaryote:dbCAN_4:dbCAN-sub-All:dbCAN_4:dbCAN-sub | -0.00772 | -0.04442 | 0.028986 | 1 |
| Eukaryote:dbCAN_4:dbCAN-sub-Bacteria:dbCAN_4:dbCAN-sub | -0.01543 | -0.05781 | 0.026948 | 0.999998 |
| Bacteria:dbCAN_4:DIAMOND-All:dbCAN_4:DIAMOND | 0.009509 | -0.02719 | 0.046211 | 1 |
| Eukaryote:dbCAN_4:DIAMOND-All:dbCAN_4:DIAMOND | -0.00951 | -0.04621 | 0.027192 | 1 |
| Eukaryote:dbCAN_4:DIAMOND-Bacteria:dbCAN_4:DIAMOND | -0.01902 | -0.0614 | 0.023361 | 0.999643 |
| Bacteria:dbCAN_4:HMMER-All:dbCAN_4:HMMER | 0.006638 | -0.03006 | 0.043339 | 1 |
| Eukaryote:dbCAN_4:HMMER-All:dbCAN_4:HMMER | -0.00664 | -0.04334 | 0.030063 | 1 |
| Eukaryote:dbCAN_4:HMMER-Bacteria:dbCAN_4:HMMER | -0.01328 | -0.05565 | 0.029103 | 1 |

## SI table 12: Output of a two-way ANOVA testing for statistically difference between prediction tools and taxonomic kingdoms for sensitivity of CAZyme/non-CAZyme classification

| Term | DF | SumSq | MeanSq | Statistic | P-value |
|---|---|---|---|---|---|
| Tax_group | 2 | 0.003398 | 0.001699 | 0.199577 | 0.819093 |
| Prediction_tool | 12 | 6.224919 | 0.518743 | 60.92723 | 2.44E-126 |
| Tax_group:Prediction_tool | 24 | 0.184762 | 0.007698 | 0.904189 | 0.597042 |
| Residuals | 2041 | 17.37737 | 0.008514 | NA | NA |

## SI table 13: Output of Tukey HSD test following a two-way ANOVA testing for statistically difference between prediction tools and taxonomic kingdoms for sensitivity of CAZyme/non-CAZyme classification

Reporting the output of a Tukey HSD test following a two-way ANOVA testing for statistically difference between prediction tools and taxonomic kingdoms for sensitivity of CAZyme/non-CAZyme classification. Specifically reporting the mean difference (MeanDiff), lower and upper 95% confidence interval (CI), and the adjusted p-value (P.Adj). Additionally, only the results comparing the output of the same tool for different taxonomic kingdoms are presented. The comprehensive output with all pairwise comparisons is available in the online repository.

| Kingdom:Classifier 1 - Kingdom:Classifier 2 | MeanDiff | Lower CI | Upper CI | P.Adj |
|---|---|---|---|---|
| Bacteria:CUPP-All:CUPP | 0.007875 | -0.06148 | 0.077233 | 1 |
| Eukaryote:CUPP-All:CUPP | -0.00787 | -0.07723 | 0.061483 | 1 |
| Eukaryote:CUPP-Bacteria:CUPP | -0.01575 | -0.09584 | 0.064338 | 1 |
| Bacteria:dbCAN_2-All:dbCAN_2 | 0.004 | -0.06536 | 0.073358 | 1 |
| Eukaryote:dbCAN_2-All:dbCAN_2 | -0.004 | -0.07336 | 0.065358 | 1 |
| Eukaryote:dbCAN_2-Bacteria:dbCAN_2 | -0.008 | -0.08809 | 0.072088 | 1 |
| Bacteria:dbCAN_2:DIAMOND-All:dbCAN_2:DIAMOND | 0.0025 | -0.06686 | 0.071858 | 1 |
| Eukaryote:dbCAN_2:DIAMOND-All:dbCAN_2:DIAMOND | -0.0025 | -0.07186 | 0.066858 | 1 |
| Eukaryote:dbCAN_2:DIAMOND-Bacteria:dbCAN_2:DIAMOND | -0.005 | -0.08509 | 0.075088 | 1 |
| Bacteria:dbCAN_2:HMMER-All:dbCAN_2:HMMER | -0.00837 | -0.07773 | 0.060983 | 1 |
| Eukaryote:dbCAN_2:HMMER-All:dbCAN_2:HMMER | 0.008375 | -0.06098 | 0.077733 | 1 |
| Eukaryote:dbCAN_2:HMMER-Bacteria:dbCAN_2:HMMER | 0.01675 | -0.06334 | 0.096838 | 1 |
| Bacteria:dbCAN_2:Hotpep-All:dbCAN_2:Hotpep | 0.011125 | -0.05823 | 0.080483 | 1 |
| Eukaryote:dbCAN_2:Hotpep-All:dbCAN_2:Hotpep | -0.01112 | -0.08048 | 0.058233 | 1 |
| Eukaryote:dbCAN_2:Hotpep-Bacteria:dbCAN_2:Hotpep | -0.02225 | -0.10234 | 0.057838 | 1 |
| Bacteria:dbCAN_3-All:dbCAN_3 | -0.00387 | -0.07323 | 0.065483 | 1 |
| Eukaryote:dbCAN_3-All:dbCAN_3 | 0.003875 | -0.06548 | 0.073233 | 1 |
| Eukaryote:dbCAN_3-Bacteria:dbCAN_3 | 0.00775 | -0.07234 | 0.087838 | 1 |
| Bacteria:dbCAN_3:DIAMOND-All:dbCAN_3:DIAMOND | -0.01287 | -0.08223 | 0.056483 | 1 |
| Eukaryote:dbCAN_3:DIAMOND-All:dbCAN_3:DIAMOND | 0.012875 | -0.05648 | 0.082233 | 1 |
| Eukaryote:dbCAN_3:DIAMOND-Bacteria:dbCAN_3:DIAMOND | 0.02575 | -0.05434 | 0.105838 | 1 |
| Bacteria:dbCAN_3:eCAMI-All:dbCAN_3:eCAMI | 0.031625 | -0.03773 | 0.100983 | 0.999501 |
| Eukaryote:dbCAN_3:eCAMI-All:dbCAN_3:eCAMI | -0.03163 | -0.10098 | 0.037733 | 0.999501 |
| Eukaryote:dbCAN_3:eCAMI-Bacteria:dbCAN_3:eCAMI | -0.06325 | -0.14334 | 0.016838 | 0.451913 |
| Bacteria:dbCAN_3:HMMER-All:dbCAN_3:HMMER | -0.01525 | -0.08461 | 0.054108 | 1 |
| Eukaryote:dbCAN_3:HMMER-All:dbCAN_3:HMMER | 0.01525 | -0.05411 | 0.084608 | 1 |
| Eukaryote:dbCAN_3:HMMER-Bacteria:dbCAN_3:HMMER | 0.0305 | -0.04959 | 0.110588 | 0.999992 |
| Bacteria:dbCAN_4-All:dbCAN_4 | -0.00937 | -0.07873 | 0.059983 | 1 |
| Eukaryote:dbCAN_4-All:dbCAN_4 | 0.009375 | -0.05998 | 0.078733 | 1 |
| Eukaryote:dbCAN_4-Bacteria:dbCAN_4 | 0.01875 | -0.06134 | 0.098838 | 1 |
| Bacteria:dbCAN_4:dbCAN-sub-All:dbCAN_4:dbCAN-sub | 0.00225 | -0.06711 | 0.071608 | 1 |
| Eukaryote:dbCAN_4:dbCAN-sub-All:dbCAN_4:dbCAN-sub | -0.00225 | -0.07161 | 0.067108 | 1 |
| Eukaryote:dbCAN_4:dbCAN-sub-Bacteria:dbCAN_4:dbCAN-sub | -0.0045 | -0.08459 | 0.075588 | 1 |
| Bacteria:dbCAN_4:DIAMOND-All:dbCAN_4:DIAMOND | -0.01963 | -0.08898 | 0.049733 | 1 |
| Eukaryote:dbCAN_4:DIAMOND-All:dbCAN_4:DIAMOND | 0.019625 | -0.04973 | 0.088983 | 1 |
| Eukaryote:dbCAN_4:DIAMOND-Bacteria:dbCAN_4:DIAMOND | 0.03925 | -0.04084 | 0.119338 | 0.997898 |
| Bacteria:dbCAN_4:HMMER-All:dbCAN_4:HMMER | -0.0135 | -0.08286 | 0.055858 | 1 |
| Eukaryote:dbCAN_4:HMMER-All:dbCAN_4:HMMER | 0.0135 | -0.05586 | 0.082858 | 1 |
| Eukaryote:dbCAN_4:HMMER-Bacteria:dbCAN_4:HMMER | 0.027 | -0.05309 | 0.107088 | 1 |

## SI table 14: Output of a two-way ANOVA testing for statistically difference between prediction tools and taxonomic kingdoms for F1-score of CAZyme/non-CAZyme classification

| Term | DF | SumSq | MeanSq | Statistic | P-value |
|---|---|---|---|---|---|
| Tax_group | 2 | 0.001974 | 0.000987 | 0.204541 | 0.815038 |
| Prediction_tool | 12 | 1.738446 | 0.14487 | 30.02434 | 6.53E-64 |
| Tax_group:Prediction_tool | 24 | 0.055394 | 0.002308 | 0.478346 | 0.984888 |
| Residuals | 2041 | 9.848032 | 0.004825 | NA | NA |

## SI table 15: Output of Tukey HSD test following a two-way ANOVA testing for statistically difference between prediction tools and taxonomic kingdoms for F1-score of CAZyme/non-CAZyme classification

Reporting the output of a Tukey HSD test following a two-way ANOVA testing for statistically difference between prediction tools and taxonomic kingdoms for F1-score of CAZyme/non-CAZyme classification. Specifically reporting the mean difference (MeanDiff), lower and upper 95% confidence interval (CI), and the adjusted p-value (P.Adj). Additionally, only the results comparing the output of the same tool for different taxonomic kingdoms are presented. The comprehensive output with all pairwise comparisons is available in the online repository.

| Kingdom:Classifier 1 - Kingdom:Classifier 2 | MeanDiff | Lower CI | Upper CI | P.Adj |
|---|---|---|---|---|
| Bacteria:CUPP-All:CUPP | 0.007415 | -0.0448 | 0.059627 | 1 |
| Eukaryote:CUPP-All:CUPP | -0.00741 | -0.05963 | 0.044798 | 1 |
| Eukaryote:CUPP-Bacteria:CUPP | -0.01483 | -0.07512 | 0.045461 | 1 |
| Bacteria:dbCAN_2-All:dbCAN_2 | 0.004046 | -0.04817 | 0.056259 | 1 |
| Eukaryote:dbCAN_2-All:dbCAN_2 | -0.00405 | -0.05626 | 0.048167 | 1 |
| Eukaryote:dbCAN_2-Bacteria:dbCAN_2 | -0.00809 | -0.06838 | 0.052197 | 1 |
| Bacteria:dbCAN_2:DIAMOND-All:dbCAN_2:DIAMOND | 0.002624 | -0.04959 | 0.054837 | 1 |
| Eukaryote:dbCAN_2:DIAMOND-All:dbCAN_2:DIAMOND | -0.00262 | -0.05484 | 0.049589 | 1 |
| Eukaryote:dbCAN_2:DIAMOND-Bacteria:dbCAN_2:DIAMOND | -0.00525 | -0.06554 | 0.055041 | 1 |
| Bacteria:dbCAN_2:HMMER-All:dbCAN_2:HMMER | -0.00375 | -0.05596 | 0.048462 | 1 |
| Eukaryote:dbCAN_2:HMMER-All:dbCAN_2:HMMER | 0.003751 | -0.04846 | 0.055964 | 1 |
| Eukaryote:dbCAN_2:HMMER-Bacteria:dbCAN_2:HMMER | 0.007501 | -0.05279 | 0.067792 | 1 |
| Bacteria:dbCAN_2:Hotpep-All:dbCAN_2:Hotpep | 0.009034 | -0.04318 | 0.061247 | 1 |
| Eukaryote:dbCAN_2:Hotpep-All:dbCAN_2:Hotpep | -0.00903 | -0.06125 | 0.043179 | 1 |
| Eukaryote:dbCAN_2:Hotpep-Bacteria:dbCAN_2:Hotpep | -0.01807 | -0.07836 | 0.042223 | 1 |
| Bacteria:dbCAN_3-All:dbCAN_3 | 0.001147 | -0.05107 | 0.05336 | 1 |
| Eukaryote:dbCAN_3-All:dbCAN_3 | -0.00115 | -0.05336 | 0.051066 | 1 |
| Eukaryote:dbCAN_3-Bacteria:dbCAN_3 | -0.00229 | -0.06258 | 0.057996 | 1 |
| Bacteria:dbCAN_3:DIAMOND-All:dbCAN_3:DIAMOND | -0.0036 | -0.05581 | 0.048613 | 1 |
| Eukaryote:dbCAN_3:DIAMOND-All:dbCAN_3:DIAMOND | 0.0036 | -0.04861 | 0.055813 | 1 |
| Eukaryote:dbCAN_3:DIAMOND-Bacteria:dbCAN_3:DIAMOND | 0.0072 | -0.05309 | 0.06749 | 1 |
| Bacteria:dbCAN_3:eCAMI-All:dbCAN_3:eCAMI | 0.019315 | -0.0329 | 0.071528 | 0.999996 |
| Eukaryote:dbCAN_3:eCAMI-All:dbCAN_3:eCAMI | -0.01932 | -0.07153 | 0.032898 | 0.999996 |
| Eukaryote:dbCAN_3:eCAMI-Bacteria:dbCAN_3:eCAMI | -0.03863 | -0.09892 | 0.02166 | 0.886832 |
| Bacteria:dbCAN_3:HMMER-All:dbCAN_3:HMMER | -0.00719 | -0.05941 | 0.045019 | 1 |
| Eukaryote:dbCAN_3:HMMER-All:dbCAN_3:HMMER | 0.007194 | -0.04502 | 0.059407 | 1 |
| Eukaryote:dbCAN_3:HMMER-Bacteria:dbCAN_3:HMMER | 0.014388 | -0.0459 | 0.074679 | 1 |
| Bacteria:dbCAN_4-All:dbCAN_4 | -0.00127 | -0.05348 | 0.050943 | 1 |
| Eukaryote:dbCAN_4-All:dbCAN_4 | 0.00127 | -0.05094 | 0.053483 | 1 |
| Eukaryote:dbCAN_4-Bacteria:dbCAN_4 | 0.002539 | -0.05775 | 0.062829 | 1 |
| Bacteria:dbCAN_4:dbCAN-sub-All:dbCAN_4:dbCAN-sub | 0.003528 | -0.04868 | 0.055741 | 1 |
| Eukaryote:dbCAN_4:dbCAN-sub-All:dbCAN_4:dbCAN-sub | -0.00353 | -0.05574 | 0.048685 | 1 |
| Eukaryote:dbCAN_4:dbCAN-sub-Bacteria:dbCAN_4:dbCAN-sub | -0.00706 | -0.06735 | 0.053234 | 1 |
| Bacteria:dbCAN_4:DIAMOND-All:dbCAN_4:DIAMOND | -0.00715 | -0.05936 | 0.045064 | 1 |
| Eukaryote:dbCAN_4:DIAMOND-All:dbCAN_4:DIAMOND | 0.007149 | -0.04506 | 0.059362 | 1 |
| Eukaryote:dbCAN_4:DIAMOND-Bacteria:dbCAN_4:DIAMOND | 0.014298 | -0.04599 | 0.074589 | 1 |
| Bacteria:dbCAN_4:HMMER-All:dbCAN_4:HMMER | -0.00624 | -0.05845 | 0.045976 | 1 |
| Eukaryote:dbCAN_4:HMMER-All:dbCAN_4:HMMER | 0.006237 | -0.04598 | 0.05845 | 1 |
| Eukaryote:dbCAN_4:HMMER-Bacteria:dbCAN_4:HMMER | 0.012474 | -0.04782 | 0.072764 | 1 |

## SI table 16: Output of a two-way ANOVA testing for statistically difference between prediction tools and taxonomic kingdoms for Accuracy of CAZyme/non-CAZyme classification

| Term | DF | SumSq | MeanSq | Statistic | P-value |
|---|---|---|---|---|---|
| Tax_group | 2 | 0.009092 | 0.004546 | 1.512065 | 0.220701 |
| Prediction_tool | 12 | 1.276489 | 0.106374 | 35.38174 | 2.53E-75 |
| Tax_group:Prediction_tool | 24 | 0.030603 | 0.001275 | 0.424124 | 0.993633 |
| Residuals | 2041 | 6.136201 | 0.003006 | NA | NA |

## SI table 17: Output of Tukey HSD test following a two-way ANOVA testing for statistically difference between prediction tools and taxonomic kingdoms for accuracy of CAZyme/non-CAZyme classification

Reporting the output of a Tukey HSD test following a two-way ANOVA testing for statistically difference between prediction tools and taxonomic kingdoms for accuracy of CAZyme/non-CAZyme classification. Specifically reporting the mean difference (MeanDiff), lower and upper 95% confidence interval (CI), and the adjusted p-value (P.Adj). Additionally, only the results comparing the output of the same tool for different taxonomic kingdoms are presented. The comprehensive output with all pairwise comparisons is available in the online repository.

| Kingdom:Classifier 1 - Kingdom:Classifier 2 | MeanDiff | Lower CI | Upper CI | P.Adj |
|---|---|---|---|---|
| Bacteria:CUPP-All:CUPP | 0.007437 | -0.03378 | 0.048652 | 1 |
| Eukaryote:CUPP-All:CUPP | -0.00744 | -0.04865 | 0.033777 | 1 |
| Eukaryote:CUPP-Bacteria:CUPP | -0.01487 | -0.06247 | 0.032716 | 1 |
| Bacteria:dbCAN_2-All:dbCAN_2 | 0.00475 | -0.03646 | 0.045965 | 1 |
| Eukaryote:dbCAN_2-All:dbCAN_2 | -0.00475 | -0.04596 | 0.036465 | 1 |
| Eukaryote:dbCAN_2-Bacteria:dbCAN_2 | -0.0095 | -0.05709 | 0.038091 | 1 |
| Bacteria:dbCAN_2:DIAMOND-All:dbCAN_2:DIAMOND | 0.003563 | -0.03765 | 0.044777 | 1 |
| Eukaryote:dbCAN_2:DIAMOND-All:dbCAN_2:DIAMOND | -0.00356 | -0.04478 | 0.037652 | 1 |
| Eukaryote:dbCAN_2:DIAMOND-Bacteria:dbCAN_2:DIAMOND | -0.00713 | -0.05472 | 0.040466 | 1 |
| Bacteria:dbCAN_2:HMMER-All:dbCAN_2:HMMER | -0.00063 | -0.04184 | 0.04059 | 1 |
| Eukaryote:dbCAN_2:HMMER-All:dbCAN_2:HMMER | 0.000625 | -0.04059 | 0.04184 | 1 |
| Eukaryote:dbCAN_2:HMMER-Bacteria:dbCAN_2:HMMER | 0.00125 | -0.04634 | 0.048841 | 1 |
| Bacteria:dbCAN_2:Hotpep-All:dbCAN_2:Hotpep | 0.007 | -0.03421 | 0.048215 | 1 |
| Eukaryote:dbCAN_2:Hotpep-All:dbCAN_2:Hotpep | -0.007 | -0.04821 | 0.034215 | 1 |
| Eukaryote:dbCAN_2:Hotpep-Bacteria:dbCAN_2:Hotpep | -0.014 | -0.06159 | 0.033591 | 1 |
| Bacteria:dbCAN_3-All:dbCAN_3 | 0.002688 | -0.03853 | 0.043902 | 1 |
| Eukaryote:dbCAN_3-All:dbCAN_3 | -0.00269 | -0.0439 | 0.038527 | 1 |
| Eukaryote:dbCAN_3-Bacteria:dbCAN_3 | -0.00538 | -0.05297 | 0.042216 | 1 |
| Bacteria:dbCAN_3:DIAMOND-All:dbCAN_3:DIAMOND | -0.00044 | -0.04165 | 0.040777 | 1 |
| Eukaryote:dbCAN_3:DIAMOND-All:dbCAN_3:DIAMOND | 0.000437 | -0.04078 | 0.041652 | 1 |
| Eukaryote:dbCAN_3:DIAMOND-Bacteria:dbCAN_3:DIAMOND | 0.000875 | -0.04672 | 0.048466 | 1 |
| Bacteria:dbCAN_3:eCAMI-All:dbCAN_3:eCAMI | 0.016812 | -0.0244 | 0.058027 | 0.999959 |
| Eukaryote:dbCAN_3:eCAMI-All:dbCAN_3:eCAMI | -0.01681 | -0.05803 | 0.024402 | 0.999959 |
| Eukaryote:dbCAN_3:eCAMI-Bacteria:dbCAN_3:eCAMI | -0.03363 | -0.08122 | 0.013966 | 0.722577 |
| Bacteria:dbCAN_3:HMMER-All:dbCAN_3:HMMER | -0.00362 | -0.04484 | 0.03759 | 1 |
| Eukaryote:dbCAN_3:HMMER-All:dbCAN_3:HMMER | 0.003625 | -0.03759 | 0.04484 | 1 |
| Eukaryote:dbCAN_3:HMMER-Bacteria:dbCAN_3:HMMER | 0.00725 | -0.04034 | 0.054841 | 1 |
| Bacteria:dbCAN_4-All:dbCAN_4 | 0.001188 | -0.04003 | 0.042402 | 1 |
| Eukaryote:dbCAN_4-All:dbCAN_4 | -0.00119 | -0.0424 | 0.040027 | 1 |
| Eukaryote:dbCAN_4-Bacteria:dbCAN_4 | -0.00238 | -0.04997 | 0.045216 | 1 |
| Bacteria:dbCAN_4:dbCAN-sub-All:dbCAN_4:dbCAN-sub | 0.00575 | -0.03546 | 0.046965 | 1 |
| Eukaryote:dbCAN_4:dbCAN-sub-All:dbCAN_4:dbCAN-sub | -0.00575 | -0.04696 | 0.035465 | 1 |
| Eukaryote:dbCAN_4:dbCAN-sub-Bacteria:dbCAN_4:dbCAN-sub | -0.0115 | -0.05909 | 0.036091 | 1 |
| Bacteria:dbCAN_4:DIAMOND-All:dbCAN_4:DIAMOND | -0.00331 | -0.04453 | 0.037902 | 1 |
| Eukaryote:dbCAN_4:DIAMOND-All:dbCAN_4:DIAMOND | 0.003312 | -0.0379 | 0.044527 | 1 |
| Eukaryote:dbCAN_4:DIAMOND-Bacteria:dbCAN_4:DIAMOND | 0.006625 | -0.04097 | 0.054216 | 1 |
| Bacteria:dbCAN_4:HMMER-All:dbCAN_4:HMMER | -0.00275 | -0.04396 | 0.038465 | 1 |
| Eukaryote:dbCAN_4:HMMER-All:dbCAN_4:HMMER | 0.00275 | -0.03846 | 0.043965 | 1 |
| Eukaryote:dbCAN_4:HMMER-Bacteria:dbCAN_4:HMMER | 0.0055 | -0.04209 | 0.053091 | 1 |

# 4 Evaluation of CAZy class classification across all CAZy class

## SI Table 18: Summary of binary CAZy class classification (overleaf)

SI table 5 lists the statistical parameter values for the binary classification of each CAZy class across all test sets. The mean (standard deviation) and lower and upper 95% confidence intervals (CIs) were calculated for each statistic by pooling all test sets, and were calculated across all CAZy classes and per CAZy class.

| CAZy Class | Classifier | Parameter | CUPP | dbCAN_2 | dbCAN_2:HMMER | dbCAN_2:DIAMOND | dbCAN_2:Hotpep | dbCAN_3 | dbCAN_3:HMMER | dbCAN_3:DIAMOND | dbCAN_3:eCAMI | dbCAN_4 | dbCAN_4:HMMER | dbCAN_4:DIAMOND | dbCAN_4:dbCAN-sub |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| All CAZy classes | Specificity | Mean | 0.995 | 0.993 | 0.994 | 0.992 | 0.973 | 0.993 | 0.994 | 0.991 | 0.983 | 0.994 | 0.994 | 0.992 | 0.993 |
| | | Standard Devision | 0.025 | 0.026 | 0.026 | 0.027 | 0.049 | 0.026 | 0.026 | 0.029 | 0.039 | 0.027 | 0.026 | 0.029 | 0.027 |
| | | Lower CI | 0.993 | 0.991 | 0.991 | 0.989 | 0.969 | 0.991 | 0.992 | 0.988 | 0.979 | 0.991 | 0.992 | 0.989 | 0.990 |
| | | Upper CI | 0.998 | 0.996 | 0.996 | 0.995 | 0.978 | 0.996 | 0.997 | 0.994 | 0.987 | 0.996 | 0.997 | 0.994 | 0.996 |
| | Sensitivity | Mean | 0.714 | 0.873 | 0.807 | 0.873 | 0.802 | 0.919 | 0.825 | 0.950 | 0.811 | 0.935 | 0.843 | 0.958 | 0.937 |
| | | Standard Devision | 0.374 | 0.173 | 0.246 | 0.204 | 0.205 | 0.145 | 0.244 | 0.128 | 0.215 | 0.129 | 0.228 | 0.111 | 0.126 |
| | | Lower CI | 0.678 | 0.857 | 0.784 | 0.853 | 0.782 | 0.905 | 0.802 | 0.937 | 0.790 | 0.923 | 0.821 | 0.947 | 0.925 |
| | | Upper CI | 0.751 | 0.890 | 0.831 | 0.893 | 0.822 | 0.933 | 0.849 | 0.962 | 0.831 | 0.948 | 0.865 | 0.969 | 0.949 |
| | Precision | Mean | 0.779 | 0.957 | 0.954 | 0.949 | 0.856 | 0.965 | 0.957 | 0.958 | 0.888 | 0.968 | 0.962 | 0.963 | 0.964 |
| | | Standard Devision | 0.398 | 0.129 | 0.143 | 0.157 | 0.244 | 0.110 | 0.135 | 0.125 | 0.218 | 0.094 | 0.124 | 0.115 | 0.100 |
| | | Lower CI | 0.740 | 0.944 | 0.941 | 0.933 | 0.832 | 0.955 | 0.944 | 0.946 | 0.867 | 0.959 | 0.950 | 0.952 | 0.955 |
| | | Upper CI | 0.817 | 0.969 | 0.968 | 0.964 | 0.880 | 0.976 | 0.970 | 0.971 | 0.909 | 0.978 | 0.974 | 0.974 | 0.974 |
| | F1-score Mean | Mean | 0.739 | 0.903 | 0.851 | 0.896 | 0.806 | 0.932 | 0.863 | 0.946 | 0.832 | 0.943 | 0.877 | 0.953 | 0.942 |
| | | Standard Devision | 0.380 | 0.144 | 0.204 | 0.176 | 0.206 | 0.120 | 0.200 | 0.117 | 0.202 | 0.104 | 0.186 | 0.105 | 0.105 |
| | | Lower CI | 0.702 | 0.889 | 0.831 | 0.879 | 0.786 | 0.920 | 0.843 | 0.935 | 0.813 | 0.933 | 0.859 | 0.943 | 0.932 |
| | | Upper CI | 0.776 | 0.916 | 0.871 | 0.913 | 0.826 | 0.944 | 0.882 | 0.957 | 0.852 | 0.953 | 0.895 | 0.963 | 0.953 |
| | Accuracy | Mean | 0.954 | 0.974 | 0.965 | 0.976 | 0.938 | 0.982 | 0.967 | 0.986 | 0.955 | 0.984 | 0.969 | 0.987 | 0.983 |
| | | Standard Devision | 0.062 | 0.043 | 0.041 | 0.045 | 0.068 | 0.034 | 0.041 | 0.032 | 0.056 | 0.033 | 0.040 | 0.032 | 0.032 |
| | | Lower CI | 0.948 | 0.970 | 0.961 | 0.972 | 0.932 | 0.979 | 0.963 | 0.983 | 0.950 | 0.981 | 0.965 | 0.984 | 0.980 |
| | | Upper CI | 0.960 | 0.978 | 0.969 | 0.981 | 0.945 | 0.985 | 0.971 | 0.989 | 0.961 | 0.987 | 0.973 | 0.990 | 0.986 |
| GH | Specificity | Mean | 0.993 | 0.992 | 0.993 | 0.986 | 0.983 | 0.992 | 0.994 | 0.984 | 0.985 | 0.993 | 0.994 | 0.986 | 0.992 |
| | | Standard Devision | 0.021 | 0.025 | 0.024 | 0.028 | 0.029 | 0.023 | 0.022 | 0.029 | 0.030 | 0.024 | 0.022 | 0.028 | 0.024 |
| | | Lower CI | 0.989 | 0.986 | 0.987 | 0.980 | 0.977 | 0.987 | 0.989 | 0.978 | 0.978 | 0.987 | 0.989 | 0.980 | 0.987 |
| | | Upper CI | 0.998 | 0.997 | 0.998 | 0.992 | 0.990 | 0.997 | 0.999 | 0.991 | 0.991 | 0.998 | 0.999 | 0.993 | 0.998 |
| | Sensitivity | Mean | 0.908 | 0.935 | 0.908 | 0.938 | 0.864 | 0.957 | 0.920 | 0.976 | 0.876 | 0.950 | 0.920 | 0.974 | 0.947 |
| | | Standard Devision | 0.067 | 0.091 | 0.082 | 0.106 | 0.118 | 0.077 | 0.083 | 0.067 | 0.110 | 0.080 | 0.083 | 0.076 | 0.079 |
| | | Lower CI | 0.893 | 0.915 | 0.890 | 0.914 | 0.838 | 0.940 | 0.901 | 0.961 | 0.852 | 0.932 | 0.902 | 0.957 | 0.930 |
| | | Upper CI | 0.923 | 0.956 | 0.926 | 0.962 | 0.890 | 0.974 | 0.938 | 0.991 | 0.901 | 0.968 | 0.939 | 0.991 | 0.965 |
| | Precision | Mean | 0.991 | 0.989 | 0.990 | 0.985 | 0.980 | 0.990 | 0.991 | 0.983 | 0.983 | 0.990 | 0.991 | 0.984 | 0.990 |
| | | Standard Devision | 0.034 | 0.035 | 0.034 | 0.036 | 0.042 | 0.033 | 0.033 | 0.039 | 0.039 | 0.034 | 0.033 | 0.039 | 0.034 |
| | | Lower CI | 0.983 | 0.981 | 0.982 | 0.977 | 0.970 | 0.982 | 0.984 | 0.974 | 0.974 | 0.982 | 0.984 | 0.975 | 0.982 |
| | | Upper CI | 0.998 | 0.996 | 0.997 | 0.993 | 0.989 | 0.997 | 0.998 | 0.992 | 0.992 | 0.997 | 0.998 | 0.993 | 0.997 |
| | F1-score Mean | Mean | 0.946 | 0.958 | 0.944 | 0.956 | 0.914 | 0.971 | 0.951 | 0.978 | 0.922 | 0.967 | 0.952 | 0.976 | 0.966 |
| | | Standard Devision | 0.045 | 0.061 | 0.059 | 0.073 | 0.077 | 0.056 | 0.060 | 0.049 | 0.070 | 0.059 | 0.060 | 0.058 | 0.058 |
| | | Lower CI | 0.936 | 0.945 | 0.931 | 0.940 | 0.896 | 0.958 | 0.938 | 0.967 | 0.907 | 0.954 | 0.938 | 0.963 | 0.953 |
| | | Upper CI | 0.956 | 0.972 | 0.957 | 0.972 | 0.931 | 0.983 | 0.965 | 0.989 | 0.938 | 0.980 | 0.965 | 0.989 | 0.978 |
| | Accuracy | Mean | 0.954 | 0.966 | 0.954 | 0.964 | 0.930 | 0.977 | 0.960 | 0.981 | 0.936 | 0.974 | 0.960 | 0.981 | 0.973 |
| | | Standard Devision | 0.029 | 0.036 | 0.033 | 0.041 | 0.049 | 0.030 | 0.033 | 0.029 | 0.048 | 0.031 | 0.033 | 0.030 | 0.030 |
| | | Lower CI | 0.947 | 0.958 | 0.946 | 0.955 | 0.919 | 0.970 | 0.953 | 0.975 | 0.926 | 0.968 | 0.953 | 0.974 | 0.967 |
| | | Upper CI | 0.961 | 0.974 | 0.961 | 0.974 | 0.941 | 0.984 | 0.967 | 0.988 | 0.947 | 0.981 | 0.967 | 0.988 | 0.980 |
| GT | Specificity | Mean | 0.992 | 0.993 | 0.990 | 0.992 | 0.992 | 0.991 | 0.990 | 0.989 | 0.992 | 0.990 | 0.990 | 0.989 | 0.990 |
| | | Standard Devision | 0.046 | 0.045 | 0.049 | 0.046 | 0.042 | 0.047 | 0.049 | 0.049 | 0.044 | 0.049 | 0.049 | 0.049 | 0.049 |
| | | Lower CI | 0.982 | 0.983 | 0.980 | 0.982 | 0.983 | 0.981 | 0.980 | 0.978 | 0.982 | 0.979 | 0.980 | 0.978 | 0.979 |
| | | Upper CI | 1.002 | 1.003 | 1.001 | 1.002 | 1.002 | 1.002 | 1.001 | 1.000 | 1.002 | 1.001 | 1.001 | 1.000 | 1.001 |
| | Sensitivity | Mean | 0.854 | 0.884 | 0.863 | 0.926 | 0.725 | 0.942 | 0.865 | 0.977 | 0.850 | 0.958 | 0.866 | 0.975 | 0.954 |
| | | Standard Devision | 0.111 | 0.138 | 0.113 | 0.151 | 0.181 | 0.097 | 0.111 | 0.090 | 0.152 | 0.092 | 0.111 | 0.091 | 0.083 |
| | | Lower CI | 0.829 | 0.854 | 0.838 | 0.892 | 0.685 | 0.920 | 0.841 | 0.957 | 0.816 | 0.937 | 0.841 | 0.955 | 0.935 |
| | | Upper CI | 0.878 | 0.915 | 0.888 | 0.959 | 0.766 | 0.964 | 0.890 | 0.997 | 0.884 | 0.978 | 0.890 | 0.995 | 0.972 |
| | Precision | Mean | 0.988 | 0.990 | 0.988 | 0.989 | 0.984 | 0.989 | 0.988 | 0.984 | 0.988 | 0.987 | 0.988 | 0.984 | 0.987 |
| | | Standard Devision | 0.062 | 0.054 | 0.057 | 0.057 | 0.069 | 0.056 | 0.057 | 0.062 | 0.057 | 0.057 | 0.057 | 0.064 | 0.057 |
| | | Lower CI | 0.974 | 0.978 | 0.976 | 0.976 | 0.968 | 0.977 | 0.976 | 0.970 | 0.975 | 0.974 | 0.976 | 0.970 | 0.974 |
| | | Upper CI | 1.002 | 1.002 | 1.001 | 1.001 | 0.999 | 1.002 | 1.001 | 0.998 | 1.001 | 0.999 | 1.001 | 0.998 | 0.999 |
| | F1-score Mean | Mean | 0.911 | 0.926 | 0.915 | 0.946 | 0.821 | 0.961 | 0.917 | 0.976 | 0.905 | 0.968 | 0.917 | 0.975 | 0.966 |
| | | Standard Devision | 0.076 | 0.102 | 0.085 | 0.112 | 0.138 | 0.079 | 0.084 | 0.076 | 0.108 | 0.075 | 0.084 | 0.078 | 0.066 |
| | | Lower CI | 0.894 | 0.903 | 0.896 | 0.921 | 0.790 | 0.943 | 0.898 | 0.960 | 0.881 | 0.951 | 0.898 | 0.958 | 0.952 |
| | | Upper CI | 0.928 | 0.948 | 0.934 | 0.971 | 0.852 | 0.978 | 0.936 | 0.993 | 0.929 | 0.985 | 0.936 | 0.992 | 0.981 |
| | Accuracy | Mean | 0.946 | 0.955 | 0.950 | 0.967 | 0.903 | 0.975 | 0.950 | 0.985 | 0.941 | 0.978 | 0.950 | 0.984 | 0.977 |
| | | Standard Devision | 0.058 | 0.077 | 0.060 | 0.082 | 0.091 | 0.059 | 0.060 | 0.057 | 0.082 | 0.057 | 0.060 | 0.058 | 0.052 |
| | | Lower CI | 0.933 | 0.938 | 0.936 | 0.949 | 0.883 | 0.962 | 0.937 | 0.972 | 0.923 | 0.965 | 0.937 | 0.971 | 0.966 |
| | | Upper CI | 0.959 | 0.972 | 0.963 | 0.985 | 0.923 | 0.988 | 0.964 | 0.997 | 0.960 | 0.991 | 0.964 | 0.997 | 0.989 |
| PL | Specificity | Mean | 1.000 | 1.000 | 1.000 | 1.000 | 0.999 | 0.999 | 0.999 | 0.999 | 1.000 | 0.999 | 0.999 | 0.999 | 0.999 |
| | | Standard Devision | 0.002 | 0.001 | 0.001 | 0.002 | 0.003 | 0.003 | 0.003 | 0.004 | 0.002 | 0.002 | 0.003 | 0.004 | 0.002 |
| | | Lower CI | 0.999 | 0.999 | 0.999 | 0.999 | 0.999 | 0.999 | 0.999 | 0.998 | 0.999 | 0.999 | 0.999 | 0.998 | 0.999 |
| | | Upper CI | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 |
| | Sensitivity | Mean | 0.851 | 0.880 | 0.897 | 0.869 | 0.841 | 0.988 | 0.974 | 0.988 | 0.796 | 0.974 | 0.974 | 0.999 | 0.975 |
| | | Standard Devision | 0.259 | 0.242 | 0.212 | 0.269 | 0.258 | 0.073 | 0.098 | 0.073 | 0.278 | 0.098 | 0.098 | 0.009 | 0.098 |
| | | Lower CI | 0.775 | 0.809 | 0.835 | 0.790 | 0.765 | 0.967 | 0.945 | 0.967 | 0.715 | 0.945 | 0.945 | 0.996 | 0.946 |
| | | Upper CI | 0.927 | 0.951 | 0.960 | 0.948 | 0.917 | 1.010 | 1.003 | 1.010 | 0.877 | 1.003 | 1.003 | 1.001 | 1.004 |
| | Precision | Mean | 0.950 | 0.953 | 0.974 | 0.925 | 0.951 | 0.985 | 0.985 | 0.973 | 0.933 | 0.985 | 0.985 | 0.973 | 0.985 |
| | | Standard Devision | 0.206 | 0.205 | 0.148 | 0.250 | 0.205 | 0.077 | 0.077 | 0.101 | 0.245 | 0.077 | 0.077 | 0.101 | 0.077 |
| | | Lower CI | 0.889 | 0.893 | 0.931 | 0.851 | 0.890 | 0.962 | 0.962 | 0.943 | 0.862 | 0.962 | 0.962 | 0.943 | 0.962 |
| | | Upper CI | 1.010 | 1.013 | 1.018 | 0.998 | 1.011 | 1.007 | 1.007 | 1.003 | 1.005 | 1.007 | 1.007 | 1.003 | 1.007 |
| | F1-score Mean | Mean | 0.885 | 0.907 | 0.925 | 0.889 | 0.880 | 0.983 | 0.974 | 0.975 | 0.847 | 0.974 | 0.974 | 0.982 | 0.975 |
| | | Standard Devision | 0.229 | 0.218 | 0.179 | 0.253 | 0.226 | 0.069 | 0.078 | 0.079 | 0.253 | 0.078 | 0.078 | 0.065 | 0.078 |
| | | Lower CI | 0.818 | 0.843 | 0.873 | 0.815 | 0.814 | 0.962 | 0.951 | 0.952 | 0.774 | 0.951 | 0.951 | 0.963 | 0.952 |
| | | Upper CI | 0.952 | 0.971 | 0.978 | 0.963 | 0.947 | 1.003 | 0.997 | 0.999 | 0.921 | 0.997 | 0.997 | 1.001 | 0.998 |
| | Accuracy | Mean | 0.994 | 0.996 | 0.996 | 0.995 | 0.993 | 0.999 | 0.999 | 0.999 | 0.992 | 0.999 | 0.999 | 0.999 | 0.999 |
| | | Standard Devision | 0.010 | 0.007 | 0.006 | 0.007 | 0.013 | 0.003 | 0.003 | 0.004 | 0.013 | 0.003 | 0.003 | 0.004 | 0.003 |
| | | Lower CI | 0.991 | 0.994 | 0.994 | 0.993 | 0.989 | 0.998 | 0.998 | 0.997 | 0.988 | 0.998 | 0.998 | 0.998 | 0.998 |
| | | Upper CI | 0.997 | 0.998 | 0.998 | 0.998 | 0.997 | 1.000 | 1.000 | 1.000 | 0.996 | 1.000 | 1.000 | 1.000 | 1.000 |
| CE | Specificity | Mean | 0.995 | 0.994 | 0.994 | 0.994 | 0.990 | 0.994 | 0.995 | 0.993 | 0.993 | 0.995 | 0.995 | 0.993 | 0.995 |
| | | Standard Devision | 0.018 | 0.022 | 0.019 | 0.021 | 0.023 | 0.022 | 0.018 | 0.023 | 0.021 | 0.018 | 0.018 | 0.023 | 0.018 |
| | | Lower CI | 0.991 | 0.989 | 0.990 | 0.989 | 0.985 | 0.989 | 0.991 | 0.987 | 0.988 | 0.991 | 0.991 | 0.988 | 0.991 |
| | | Upper CI | 1.000 | 0.999 | 0.999 | 0.999 | 0.996 | 0.999 | 0.999 | 0.998 | 0.997 | 0.999 | 0.999 | 0.998 | 0.999 |
| | Sensitivity | Mean | 0.911 | 0.921 | 0.921 | 0.848 | 0.851 | 0.928 | 0.923 | 0.930 | 0.807 | 0.978 | 0.953 | 0.975 | 0.975 |
| | | Standard Devision | 0.133 | 0.143 | 0.136 | 0.247 | 0.206 | 0.154 | 0.133 | 0.170 | 0.245 | 0.067 | 0.101 | 0.085 | 0.068 |
| | | Lower CI | 0.881 | 0.889 | 0.890 | 0.792 | 0.804 | 0.893 | 0.892 | 0.891 | 0.751 | 0.963 | 0.930 | 0.956 | 0.960 |
| | | Upper CI | 0.942 | 0.954 | 0.952 | 0.904 | 0.898 | 0.963 | 0.953 | 0.969 | 0.863 | 0.994 | 0.976 | 0.995 | 0.991 |
| | Precision | Mean | 0.960 | 0.952 | 0.953 | 0.928 | 0.913 | 0.953 | 0.960 | 0.930 | 0.916 | 0.960 | 0.960 | 0.948 | 0.959 |
| | | Standard Devision | 0.113 | 0.141 | 0.119 | 0.209 | 0.167 | 0.139 | 0.110 | 0.186 | 0.191 | 0.110 | 0.110 | 0.151 | 0.112 |
| | | Lower CI | 0.934 | 0.920 | 0.925 | 0.880 | 0.875 | 0.921 | 0.935 | 0.888 | 0.872 | 0.935 | 0.935 | 0.913 | 0.934 |
| | | Upper CI | 0.986 | 0.984 | 0.980 | 0.976 | 0.951 | 0.984 | 0.985 | 0.973 | 0.959 | 0.986 | 0.986 | 0.982 | 0.985 |
| | F1-score Mean | Mean | 0.925 | 0.922 | 0.925 | 0.864 | 0.855 | 0.924 | 0.930 | 0.915 | 0.831 | 0.964 | 0.949 | 0.951 | 0.962 |
| | | Standard Devision | 0.109 | 0.130 | 0.112 | 0.223 | 0.171 | 0.140 | 0.107 | 0.168 | 0.202 | 0.085 | 0.096 | 0.119 | 0.086 |
| | | Lower CI | 0.900 | 0.893 | 0.899 | 0.813 | 0.816 | 0.892 | 0.906 | 0.877 | 0.785 | 0.944 | 0.927 | 0.923 | 0.942 |
| | | Upper CI | 0.950 | 0.952 | 0.950 | 0.915 | 0.894 | 0.956 | 0.955 | 0.954 | 0.878 | 0.983 | 0.971 | 0.978 | 0.981 |
| | Accuracy | Mean | 0.990 | 0.989 | 0.990 | 0.986 | 0.982 | 0.990 | 0.991 | 0.989 | 0.982 | 0.994 | 0.993 | 0.992 | 0.994 |
| | | Standard Devision | 0.019 | 0.023 | 0.019 | 0.024 | 0.025 | 0.022 | 0.018 | 0.023 | 0.025 | 0.017 | 0.017 | 0.022 | 0.017 |
| | | Lower CI | 0.986 | 0.984 | 0.986 | 0.981 | 0.976 | 0.985 | 0.987 | 0.984 | 0.976 | 0.990 | 0.989 | 0.987 | 0.990 |
| | | Upper CI | 0.994 | 0.994 | 0.994 | 0.992 | 0.988 | 0.995 | 0.994 | 0.994 | 0.987 | 0.998 | 0.997 | 0.997 | 0.998 |
| AA | Specificity | Mean | 0.992 | 0.992 | 0.991 | 0.992 | 0.992 | 0.991 | 0.991 | 0.990 | 0.992 | 0.991 | 0.991 | 0.990 | 0.990 |
| | | Standard Devision | 0.017 | 0.018 | 0.019 | 0.018 | 0.018 | 0.020 | 0.020 | 0.022 | 0.017 | 0.020 | 0.020 | 0.022 | 0.020 |
| | | Lower CI | 0.987 | 0.987 | 0.986 | 0.987 | 0.987 | 0.985 | 0.985 | 0.984 | 0.987 | 0.985 | 0.985 | 0.984 | 0.985 |
| | | Upper CI | 0.997 | 0.997 | 0.997 | 0.997 | 0.997 | 0.997 | 0.997 | 0.996 | 0.997 | 0.997 | 0.997 | 0.997 | 0.996 |
| | Sensitivity | Mean | 0.905 | 0.919 | 0.937 | 0.869 | 0.874 | 0.984 | 0.988 | 0.988 | 0.782 | 0.994 | 0.988 | 0.970 | 0.989 |
| | | Standard Devision | 0.130 | 0.115 | 0.092 | 0.179 | 0.194 | 0.045 | 0.039 | 0.036 | 0.261 | 0.026 | 0.039 | 0.095 | 0.033 |
| | | Lower CI | 0.867 | 0.886 | 0.911 | 0.817 | 0.817 | 0.970 | 0.977 | 0.978 | 0.706 | 0.986 | 0.977 | 0.942 | 0.980 |
| | | Upper CI | 0.942 | 0.952 | 0.964 | 0.921 | 0.930 | 0.997 | 0.999 | 0.999 | 0.858 | 1.001 | 0.999 | 0.998 | 0.999 |
| | Precision | Mean | 0.918 | 0.921 | 0.919 | 0.923 | 0.898 | 0.924 | 0.919 | 0.917 | 0.851 | 0.920 | 0.919 | 0.920 | 0.918 |
| | | Standard Devision | 0.170 | 0.169 | 0.161 | 0.171 | 0.217 | 0.161 | 0.162 | 0.169 | 0.281 | 0.161 | 0.162 | 0.170 | 0.162 |
| | | Lower CI | 0.869 | 0.872 | 0.872 | 0.873 | 0.835 | 0.877 | 0.872 | 0.868 | 0.770 | 0.873 | 0.872 | 0.870 | 0.871 |
| | | Upper CI | 0.968 | 0.970 | 0.966 | 0.972 | 0.961 | 0.970 | 0.966 | 0.966 | 0.933 | 0.967 | 0.966 | 0.969 | 0.965 |
| | F1-score Mean | Mean | 0.901 | 0.910 | 0.917 | 0.877 | 0.873 | 0.944 | 0.943 | 0.941 | 0.801 | 0.946 | 0.943 | 0.931 | 0.943 |
| | | Standard Devision | 0.138 | 0.132 | 0.114 | 0.160 | 0.192 | 0.110 | 0.111 | 0.114 | 0.256 | 0.109 | 0.111 | 0.123 | 0.109 |
| | | Lower CI | 0.861 | 0.872 | 0.884 | 0.830 | 0.817 | 0.912 | 0.911 | 0.908 | 0.726 | 0.915 | 0.911 | 0.895 | 0.911 |
| | | Upper CI | 0.941 | 0.949 | 0.950 | 0.923 | 0.929 | 0.976 | 0.976 | 0.974 | 0.875 | 0.978 | 0.976 | 0.966 | 0.975 |
| | Accuracy | Mean | 0.984 | 0.986 | 0.986 | 0.984 | 0.983 | 0.990 | 0.990 | 0.990 | 0.979 | 0.991 | 0.990 | 0.990 | 0.990 |
| | | Standard Devision | 0.022 | 0.020 | 0.020 | 0.021 | 0.023 | 0.019 | 0.020 | 0.020 | 0.023 | 0.019 | 0.020 | 0.020 | 0.019 |
| | | Lower CI | 0.978 | 0.980 | 0.980 | 0.978 | 0.976 | 0.985 | 0.984 | 0.984 | 0.972 | 0.985 | 0.984 | 0.984 | 0.985 |
| | | Upper CI | 0.990 | 0.991 | 0.991 | 0.990 | 0.989 | 0.996 | 0.996 | 0.995 | 0.986 | 0.996 | 0.996 | 0.995 | 0.996 |
| CBM | Specificity | Mean | 1.000 | 0.992 | 0.996 | 0.991 | 0.901 | 0.994 | 0.996 | 0.993 | 0.947 | 0.995 | 0.996 | 0.994 | 0.993 |
| | | Standard Devision | 0.000 | 0.012 | 0.009 | 0.015 | 0.051 | 0.011 | 0.009 | 0.014 | 0.052 | 0.010 | 0.010 | 0.015 | 0.013 |
| | | Lower CI | 1.000 | 0.990 | 0.994 | 0.988 | 0.890 | 0.992 | 0.994 | 0.990 | 0.935 | 0.993 | 0.994 | 0.991 | 0.990 |
| | | Upper CI | 1.000 | 0.995 | 0.998 | 0.995 | 0.913 | 0.997 | 0.998 | 0.996 | 0.959 | 0.997 | 0.998 | 0.997 | 0.996 |
| | Sensitivity | Mean | 0.000 | 0.724 | 0.413 | 0.785 | 0.706 | 0.768 | 0.413 | 0.867 | 0.735 | 0.799 | 0.474 | 0.877 | 0.819 |
| | | Standard Devision | 0.000 | 0.192 | 0.217 | 0.219 | 0.210 | 0.193 | 0.215 | 0.173 | 0.220 | 0.192 | 0.236 | 0.170 | 0.200 |
| | | Lower CI | 0.000 | 0.681 | 0.365 | 0.737 | 0.659 | 0.725 | 0.366 | 0.829 | 0.686 | 0.757 | 0.422 | 0.840 | 0.775 |
| | | Upper CI | 0.000 | 0.767 | 0.461 | 0.834 | 0.753 | 0.811 | 0.461 | 0.906 | 0.784 | 0.842 | 0.527 | 0.915 | 0.864 |
| | Precision | Mean | 0.000 | 0.920 | 0.897 | 0.922 | 0.470 | 0.943 | 0.896 | 0.951 | 0.662 | 0.956 | 0.918 | 0.957 | 0.938 |
| | | Standard Devision | 0.000 | 0.133 | 0.230 | 0.150 | 0.173 | 0.136 | 0.230 | 0.128 | 0.225 | 0.092 | 0.200 | 0.127 | 0.115 |
| | | Lower CI | 0.000 | 0.891 | 0.845 | 0.889 | 0.431 | 0.913 | 0.845 | 0.923 | 0.611 | 0.936 | 0.874 | 0.929 | 0.912 |
| | | Upper CI | 0.000 | 0.950 | 0.948 | 0.955 | 0.508 | 0.973 | 0.948 | 0.980 | 0.712 | 0.977 | 0.963 | 0.985 | 0.963 |
| | F1-score Mean | Mean | 0.000 | 0.797 | 0.541 | 0.833 | 0.551 | 0.835 | 0.542 | 0.899 | 0.681 | 0.855 | 0.595 | 0.907 | 0.858 |
| | | Standard Devision | 0.000 | 0.154 | 0.215 | 0.176 | 0.172 | 0.155 | 0.213 | 0.138 | 0.208 | 0.139 | 0.220 | 0.136 | 0.149 |
| | | Lower CI | 0.000 | 0.763 | 0.494 | 0.793 | 0.512 | 0.801 | 0.495 | 0.868 | 0.635 | 0.824 | 0.546 | 0.877 | 0.824 |
| | | Upper CI | 0.000 | 0.831 | 0.589 | 0.872 | 0.589 | 0.870 | 0.590 | 0.929 | 0.727 | 0.886 | 0.644 | 0.937 | 0.891 |
| | Accuracy | Mean | 0.885 | 0.966 | 0.938 | 0.972 | 0.882 | 0.972 | 0.938 | 0.981 | 0.927 | 0.976 | 0.944 | 0.983 | 0.976 |
| | | Standard Devision | 0.085 | 0.026 | 0.030 | 0.025 | 0.055 | 0.023 | 0.030 | 0.019 | 0.056 | 0.021 | 0.030 | 0.018 | 0.024 |
| | | Lower CI | 0.866 | 0.960 | 0.931 | 0.966 | 0.870 | 0.967 | 0.931 | 0.977 | 0.914 | 0.972 | 0.937 | 0.979 | 0.971 |
| | | Upper CI | 0.904 | 0.971 | 0.944 | 0.978 | 0.895 | 0.977 | 0.945 | 0.986 | 0.939 | 0.981 | 0.950 | 0.987 | 0.982 |

## SI Figure 13: Binary CAZy class classification performance per test set

Statistical parameters were calculated for the binary classification per CAZy class for each test set. SI figure 13 plots the values of each statistical parameter per test set.

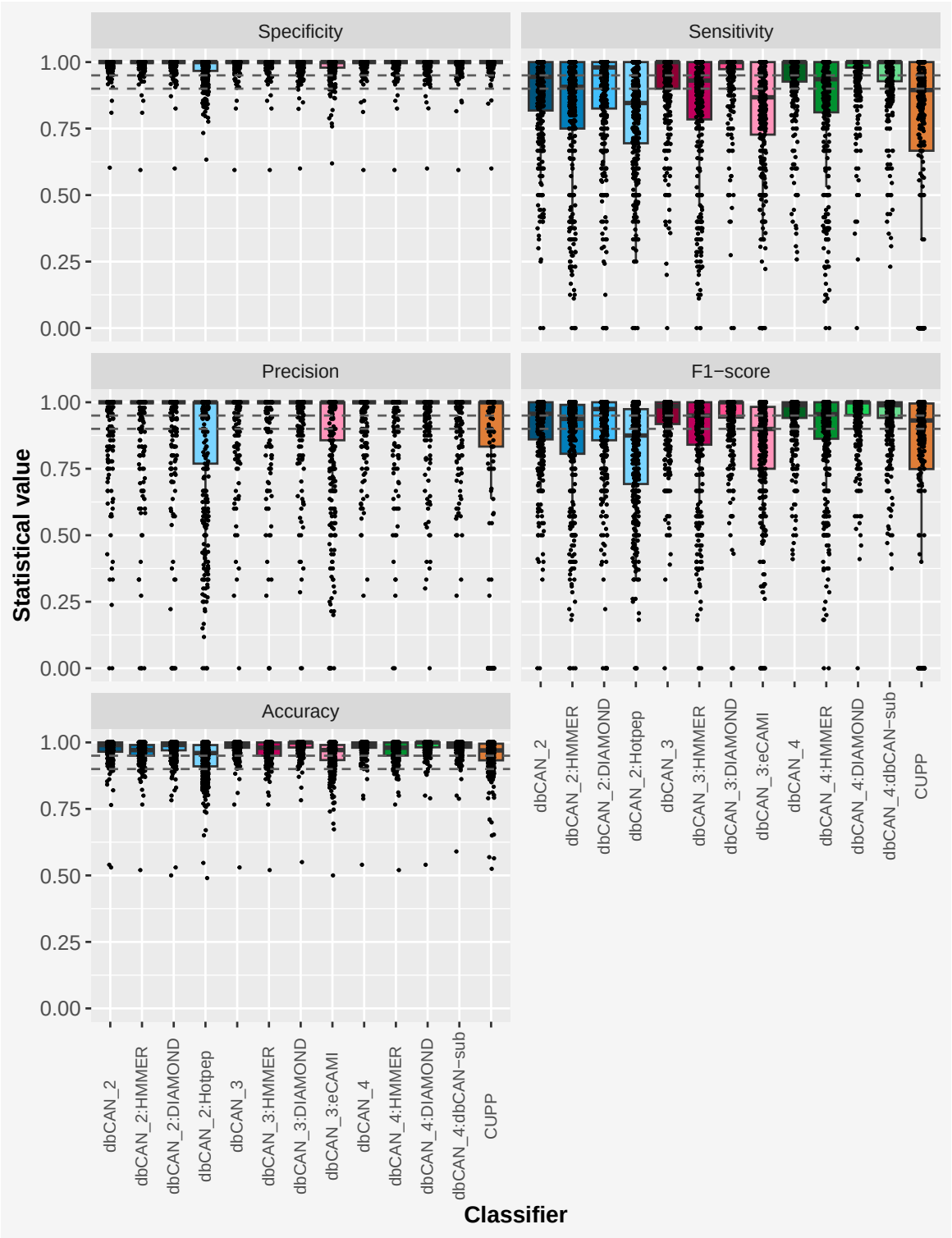



Figure 13: One-dimensional scatter plot overlaying a box and whisker plot, where each point in the scatter plot represents a test set and the corresponding value of each statistical parameter estimating the performance of binary CAZy class classification of protein sequences by CAZyme classifiers, pooling CAZy classes. Where a tool is incorporated into a CAZyme classifier, the tool name is prefixed with the respective CAZyme classifier. dbCAN_2 and associated tools are coloured blue; dbCAN_3 pink; dbCAN_4 green; and CUPP orange.

## 4.1 Statistically testing for difference between the means of overall CAZy class classification

### SI Table 19: Tukey HSD test to measure the statistically significant difference between the mean F1-scores for overall CAZy class classification (overleaf)

Tukey HSD test of the mean F1-score between the classifiers, reporting the lower and upper 95% confidence interval, and the adjusted p-value (P-Adj).

| Classifier 1 - vs - Classifier 2 | Mean Difference | Lower 95% Confidence Interval | Upper 95% Confidence Interval | Adjusted P-value |
|---|---|---|---|---|
| dbCAN_2:HMMER-dbCAN_2 | -0.051258066 | -0.094596599 | -0.007919533 | 0.00594381 |
| dbCAN_2:DIAMOND-dbCAN_2 | -0.006414152 | -0.049752685 | 0.036924381 | 0.999999445 |
| dbCAN_2:Hotpep-dbCAN_2 | -0.096851055 | -0.140189588 | -0.053512522 | 8.23E-12 |
| dbCAN_3-dbCAN_2 | 0.029370465 | -0.013968068 | 0.072708998 | 0.557038419 |
| dbCAN_3:HMMER-dbCAN_2 | -0.03969143 | -0.083029963 | 0.003647103 | 0.112264186 |
| dbCAN_3:DIAMOND-dbCAN_2 | 0.043422351 | 8.38E-05 | 0.086760884 | 0.049013102 |
| dbCAN_3:eCAMI-dbCAN_2 | -0.070214155 | -0.113526382 | -0.026901928 | 6.21E-06 |
| dbCAN_4-dbCAN_2 | 0.040593571 | -0.002744962 | 0.083932104 | 0.092914316 |
| dbCAN_4:HMMER-dbCAN_2 | -0.025895496 | -0.069234029 | 0.017443037 | 0.74604168 |
| dbCAN_4:DIAMOND-dbCAN_2 | 0.050725033 | 0.0073865 | 0.094063566 | 0.006962678 |
| dbCAN_4:dbCAN-sub-dbCAN_2 | 0.039885748 | -0.003452785 | 0.083224281 | 0.107848544 |
| CUPP-dbCAN_2 | -0.163616504 | -0.206955037 | -0.120277971 | 0 |
| dbCAN_2:DIAMOND-dbCAN_2:HMMER | 0.044843914 | 0.001505381 | 0.088182447 | 0.034643646 |
| dbCAN_2:Hotpep-dbCAN_2:HMMER | -0.045592989 | -0.088931522 | -0.002254456 | 0.02866431 |
| dbCAN_3-dbCAN_2:HMMER | 0.080628531 | 0.037289998 | 0.123967064 | 5.85E-08 |
| dbCAN_3:HMMER-dbCAN_2:HMMER | 0.011566636 | -0.031771897 | 0.054905169 | 0.999641912 |
| dbCAN_3:DIAMOND-dbCAN_2:HMMER | 0.094680417 | 0.051341884 | 0.13801895 | 3.63E-11 |
| dbCAN_3:eCAMI-dbCAN_2:HMMER | -0.018956089 | -0.062268316 | 0.024356138 | 0.966279531 |
| dbCAN_4-dbCAN_2:HMMER | 0.091851637 | 0.048513104 | 0.13519017 | 1.85E-10 |
| dbCAN_4:HMMER-dbCAN_2:HMMER | 0.02536257 | -0.017975963 | 0.068701103 | 0.771883331 |
| dbCAN_4:DIAMOND-dbCAN_2:HMMER | 0.101983099 | 0.058644566 | 0.145321632 | 0 |
| dbCAN_4:dbCAN-sub-dbCAN_2:HMMER | 0.091143814 | 0.047805281 | 0.134482347 | 2.73E-10 |
| CUPP-dbCAN_2:HMMER | -0.112358438 | -0.155696971 | -0.069019905 | 0 |
| dbCAN_2:Hotpep-dbCAN_2:DIAMOND | -0.090436903 | -0.133775436 | -0.04709837 | 4.01E-10 |
| dbCAN_3-dbCAN_2:DIAMOND | 0.035784617 | -0.007553916 | 0.07912315 | 0.233368764 |
| dbCAN_3:HMMER-dbCAN_2:DIAMOND | -0.033277278 | -0.076615811 | 0.010061255 | 0.344959813 |
| dbCAN_3:DIAMOND-dbCAN_2:DIAMOND | 0.049836503 | 0.00649797 | 0.093175036 | 0.009022462 |
| dbCAN_3:eCAMI-dbCAN_2:DIAMOND | -0.063800003 | -0.107112231 | -0.020487776 | 8.13E-05 |
| dbCAN_4-dbCAN_2:DIAMOND | 0.047007722 | 0.003669189 | 0.090346255 | 0.01979958 |
| dbCAN_4:HMMER-dbCAN_2:DIAMOND | -0.019481344 | -0.062819877 | 0.023857189 | 0.958607738 |
| dbCAN_4:DIAMOND-dbCAN_2:DIAMOND | 0.057139184 | 0.013800651 | 0.100477717 | 0.000909615 |
| dbCAN_4:dbCAN-sub-dbCAN_2:DIAMOND | 0.046299899 | 0.002961367 | 0.089638432 | 0.02387292 |
| CUPP-dbCAN_2:DIAMOND | -0.157202352 | -0.200540885 | -0.113863819 | 0 |
| dbCAN_3-dbCAN_2:Hotpep | 0.12622152 | 0.082882987 | 0.169560053 | 0 |
| dbCAN_3:HMMER-dbCAN_2:Hotpep | 0.057159625 | 0.013821092 | 0.100498158 | 0.000903332 |
| dbCAN_3:DIAMOND-dbCAN_2:Hotpep | 0.140273406 | 0.096934873 | 0.183611939 | 0 |
| dbCAN_3:eCAMI-dbCAN_2:Hotpep | 0.0266369 | -0.016675327 | 0.069949127 | 0.707391413 |
| dbCAN_4-dbCAN_2:Hotpep | 0.137444625 | 0.094106092 | 0.180783158 | 0 |
| dbCAN_4:HMMER-dbCAN_2:Hotpep | 0.070955559 | 0.027617026 | 0.114294092 | 4.62E-06 |
| dbCAN_4:DIAMOND-dbCAN_2:Hotpep | 0.147576087 | 0.104237554 | 0.19091462 | 0 |
| dbCAN_4:dbCAN-sub-dbCAN_2:Hotpep | 0.136736803 | 0.09339827 | 0.180075336 | 0 |
| CUPP-dbCAN_2:Hotpep | -0.066765449 | -0.110103982 | -0.023426916 | 2.59E-05 |
| dbCAN_3:HMMER-dbCAN_3 | -0.069061895 | -0.112400428 | -0.025723362 | 1.02E-05 |
| dbCAN_3:DIAMOND-dbCAN_3 | 0.014051886 | -0.029286647 | 0.057390419 | 0.997529557 |
| dbCAN_3:eCAMI-dbCAN_3 | -0.09958462 | -0.142896847 | -0.056272393 | 0 |
| dbCAN_4-dbCAN_3 | 0.011223105 | -0.032115428 | 0.054561638 | 0.999737731 |
| dbCAN_4:HMMER-dbCAN_3 | -0.055265961 | -0.098604494 | -0.011927428 | 0.001696742 |
| dbCAN_4:DIAMOND-dbCAN_3 | 0.021354567 | -0.021983965 | 0.0646931 | 0.919738998 |
| dbCAN_4:dbCAN-sub-dbCAN_3 | 0.010515283 | -0.03282325 | 0.053853816 | 0.999867429 |
| CUPP-dbCAN_3 | -0.192986969 | -0.236325502 | -0.149648436 | 0 |
| dbCAN_3:DIAMOND-dbCAN_3:HMMER | 0.083113781 | 0.039775248 | 0.126452314 | 1.75E-08 |
| dbCAN_3:eCAMI-dbCAN_3:HMMER | -0.030522725 | -0.073834952 | 0.012789502 | 0.490720739 |
| dbCAN_4-dbCAN_3:HMMER | 0.080285001 | 0.036946468 | 0.123623534 | 6.89E-08 |
| dbCAN_4:HMMER-dbCAN_3:HMMER | 0.013795934 | -0.029542599 | 0.057134467 | 0.997927902 |
| dbCAN_4:DIAMOND-dbCAN_3:HMMER | 0.090416463 | 0.04707793 | 0.133754996 | 4.06E-10 |
| dbCAN_4:dbCAN-sub-dbCAN_3:HMMER | 0.079577178 | 0.036238645 | 0.122915711 | 9.65E-08 |
| CUPP-dbCAN_3:HMMER | -0.123925074 | -0.167263607 | -0.080586541 | 0 |
| dbCAN_3:eCAMI-dbCAN_3:DIAMOND | -0.113636506 | -0.156948733 | -0.070324279 | 0 |
| dbCAN_4-dbCAN_3:DIAMOND | -0.00282878 | -0.046167313 | 0.040509753 | 1 |
| dbCAN_4:HMMER-dbCAN_3:DIAMOND | -0.069317847 | -0.11265638 | -0.025979314 | 9.19E-06 |
| dbCAN_4:DIAMOND-dbCAN_3:DIAMOND | 0.007302682 | -0.036035851 | 0.050641215 | 0.999997571 |
| dbCAN_4:dbCAN-sub-dbCAN_3:DIAMOND | -0.003536603 | -0.046875136 | 0.03980193 | 0.999999999 |
| CUPP-dbCAN_3:DIAMOND | -0.207038855 | -0.250377388 | -0.163700322 | 0 |
| dbCAN_4-dbCAN_3:eCAMI | 0.110807726 | 0.067495498 | 0.154119953 | 0 |
| dbCAN_4:HMMER-dbCAN_3:eCAMI | 0.044318659 | 0.001006432 | 0.087630886 | 0.039198582 |
| dbCAN_4:DIAMOND-dbCAN_3:eCAMI | 0.120939188 | 0.07762696 | 0.164251415 | 0 |
| dbCAN_4:dbCAN-sub-dbCAN_3:eCAMI | 0.110099903 | 0.066787675 | 0.15341213 | 0 |
| CUPP-dbCAN_3:eCAMI | -0.093402349 | -0.136714576 | -0.050090122 | 7.49E-11 |
| dbCAN_4:HMMER-dbCAN_4 | -0.066489066 | -0.109827599 | -0.023150533 | 2.90E-05 |
| dbCAN_4:DIAMOND-dbCAN_4 | 0.010131462 | -0.033207071 | 0.053469995 | 0.999910666 |
| dbCAN_4:dbCAN-sub-dbCAN_4 | -0.000707823 | -0.044046356 | 0.04263071 | 1 |
| CUPP-dbCAN_4 | -0.204210075 | -0.247548608 | -0.160871542 | 0 |
| dbCAN_4:DIAMOND-dbCAN_4:HMMER | 0.076620528 | 0.033281996 | 0.119959061 | 3.82E-07 |
| dbCAN_4:dbCAN-sub-dbCAN_4:HMMER | 0.065781244 | 0.022442711 | 0.109119777 | 3.83E-05 |
| CUPP-dbCAN_4:HMMER | -0.137721008 | -0.181059541 | -0.094382475 | 0 |
| dbCAN_4:dbCAN-sub-dbCAN_4:DIAMOND | -0.010839285 | -0.054177818 | 0.032499248 | 0.999817549 |
| CUPP-dbCAN_4:DIAMOND | -0.214341537 | -0.25768007 | -0.171003004 | 0 |
| CUPP-dbCAN_4:dbCAN-sub | -0.203502252 | -0.246840785 | -0.160163719 | 0 |

## SI Table 20: Tukey HSD test to measure the statistically significant difference between the mean accuracy for overall CAZy class classification (overleaf)

Tukey HSD test of the mean accuracy between the classifiers, reporting the lower and upper 95% confidence interval, and the adjusted p-value (P-Adj).

| Classifier 1 - vs - Classifier 2 | Mean Difference | Lower 95% Confidence Interval | Upper 95% Confidence Interval | Adjusted P-value |
|---|---|---|---|---|
| dbCAN_2:HMMER-dbCAN_2 | -0.008598102 | -0.018902681 | 0.001706476 | 0.218857961 |
| dbCAN_2:DIAMOND-dbCAN_2 | 0.002480242 | -0.007824337 | 0.012784821 | 0.99987821 |
| dbCAN_2:Hotpep-dbCAN_2 | -0.035347415 | -0.045651993 | -0.025042836 | 0 |
| dbCAN_3-dbCAN_2 | 0.008195934 | -0.002108644 | 0.018500513 | 0.288718371 |
| dbCAN_3:HMMER-dbCAN_2 | -0.006305164 | -0.016609743 | 0.003999415 | 0.71443295 |
| dbCAN_3:DIAMOND-dbCAN_2 | 0.01260771 | 0.002303131 | 0.022912289 | 0.003467539 |
| dbCAN_3:eCAMI-dbCAN_2 | -0.018472287 | -0.028770611 | -0.008173963 | 2.28E-07 |
| dbCAN_4-dbCAN_2 | 0.010085367 | -0.000219212 | 0.020389946 | 0.062015783 |
| dbCAN_4:HMMER-dbCAN_2 | -0.004807675 | -0.015112254 | 0.005496904 | 0.945427154 |
| dbCAN_4:DIAMOND-dbCAN_2 | 0.013296096 | 0.002991517 | 0.023600675 | 0.001368409 |
| dbCAN_4:dbCAN-sub-dbCAN_2 | 0.009522752 | -0.000781827 | 0.019827331 | 0.10421616 |
| CUPP-dbCAN_2 | -0.019920069 | -0.030224648 | -0.00961549 | 1.26E-08 |
| dbCAN_2:DIAMOND-dbCAN_2:HMMER | 0.011078344 | 0.000773765 | 0.021382923 | 0.022104988 |
| dbCAN_2:Hotpep-dbCAN_2:HMMER | -0.026749312 | -0.037053891 | -0.016444733 | 0 |
| dbCAN_3-dbCAN_2:HMMER | 0.016794037 | 0.006489458 | 0.027098616 | 5.30E-06 |
| dbCAN_3:HMMER-dbCAN_2:HMMER | 0.002292938 | -0.008011641 | 0.012597517 | 0.999947387 |
| dbCAN_3:DIAMOND-dbCAN_2:HMMER | 0.021205812 | 0.010901233 | 0.031510391 | 7.83E-10 |
| dbCAN_3:eCAMI-dbCAN_2:HMMER | -0.009874184 | -0.020172508 | 0.00042414 | 0.075368225 |
| dbCAN_4-dbCAN_2:HMMER | 0.01868347 | 0.008378891 | 0.028988048 | 1.54E-07 |
| dbCAN_4:HMMER-dbCAN_2:HMMER | 0.003790427 | -0.006514152 | 0.014095006 | 0.992098863 |
| dbCAN_4:DIAMOND-dbCAN_2:HMMER | 0.021894198 | 0.01158962 | 0.032198777 | 1.63E-10 |
| dbCAN_4:dbCAN-sub-dbCAN_2:HMMER | 0.018120854 | 0.007816275 | 0.028425433 | 4.60E-07 |
| CUPP-dbCAN_2:HMMER | -0.011321967 | -0.021626545 | -0.001017388 | 0.01680028 |
| dbCAN_2:Hotpep-dbCAN_2:DIAMOND | -0.037827656 | -0.048132235 | -0.027523078 | 0 |
| dbCAN_3-dbCAN_2:DIAMOND | 0.005715693 | -0.004588886 | 0.016020271 | 0.830349104 |
| dbCAN_3:HMMER-dbCAN_2:DIAMOND | -0.008785406 | -0.019089985 | 0.001519173 | 0.190518136 |
| dbCAN_3:DIAMOND-dbCAN_2:DIAMOND | 0.010127468 | -0.000177111 | 0.020432047 | 0.059536315 |
| dbCAN_3:eCAMI-dbCAN_2:DIAMOND | -0.020952529 | -0.031250853 | -0.010654204 | 1.33E-09 |
| dbCAN_4-dbCAN_2:DIAMOND | 0.007605125 | -0.002699453 | 0.017909704 | 0.4114252 |
| dbCAN_4:HMMER-dbCAN_2:DIAMOND | -0.007287917 | -0.017592496 | 0.003016662 | 0.484554815 |
| dbCAN_4:DIAMOND-dbCAN_2:DIAMOND | 0.010815854 | 0.000511275 | 0.021120433 | 0.02943742 |
| dbCAN_4:dbCAN-sub-dbCAN_2:DIAMOND | 0.00704251 | -0.003262069 | 0.017347089 | 0.542901279 |
| CUPP-dbCAN_2:DIAMOND | -0.022400311 | -0.032704889 | -0.012095732 | 4.83E-11 |
| dbCAN_3-dbCAN_2:Hotpep | 0.043543349 | 0.03323877 | 0.053847928 | 0 |
| dbCAN_3:HMMER-dbCAN_2:Hotpep | 0.02904225 | 0.018737672 | 0.039346829 | 0 |
| dbCAN_3:DIAMOND-dbCAN_2:Hotpep | 0.047955124 | 0.037650546 | 0.058259703 | 0 |
| dbCAN_3:eCAMI-dbCAN_2:Hotpep | 0.016875128 | 0.006576804 | 0.027173452 | 4.51E-06 |
| dbCAN_4-dbCAN_2:Hotpep | 0.045432782 | 0.035128203 | 0.055737361 | 0 |
| dbCAN_4:HMMER-dbCAN_2:Hotpep | 0.03053974 | 0.020235161 | 0.040844318 | 0 |
| dbCAN_4:DIAMOND-dbCAN_2:Hotpep | 0.048643511 | 0.038338932 | 0.05894809 | 0 |
| dbCAN_4:dbCAN-sub-dbCAN_2:Hotpep | 0.044870166 | 0.034565588 | 0.055174745 | 0 |
| CUPP-dbCAN_2:Hotpep | 0.015427346 | 0.005122767 | 0.025731925 | 5.44E-05 |
| dbCAN_3:HMMER-dbCAN_3 | -0.014501099 | -0.024805677 | -0.00419652 | 0.000235155 |
| dbCAN_3:DIAMOND-dbCAN_3 | 0.004411775 | -0.005892803 | 0.014716354 | 0.971626487 |
| dbCAN_3:eCAMI-dbCAN_3 | -0.026668221 | -0.036966545 | -0.016369897 | 0 |
| dbCAN_4-dbCAN_3 | 0.001889433 | -0.008415146 | 0.012194012 | 0.999993727 |
| dbCAN_4:HMMER-dbCAN_3 | -0.013003609 | -0.023308188 | -0.002699031 | 0.002045575 |
| dbCAN_4:DIAMOND-dbCAN_3 | 0.005100162 | -0.005204417 | 0.01540474 | 0.917259301 |
| dbCAN_4:dbCAN-sub-dbCAN_3 | 0.001326817 | -0.008977761 | 0.011631396 | 0.999999888 |
| CUPP-dbCAN_3 | -0.028116003 | -0.038420582 | -0.017811424 | 0 |
| dbCAN_3:DIAMOND-dbCAN_3:HMMER | 0.018912874 | 0.008608295 | 0.029217453 | 9.81E-08 |
| dbCAN_3:eCAMI-dbCAN_3:HMMER | -0.012167123 | -0.022465447 | -0.001868798 | 0.006042303 |
| dbCAN_4-dbCAN_3:HMMER | 0.016390531 | 0.006085952 | 0.02669511 | 1.08E-05 |
| dbCAN_4:HMMER-dbCAN_3:HMMER | 0.001497489 | -0.00880709 | 0.011802068 | 0.99999955 |
| dbCAN_4:DIAMOND-dbCAN_3:HMMER | 0.01960126 | 0.009296681 | 0.029905839 | 2.43E-08 |
| dbCAN_4:dbCAN-sub-dbCAN_3:HMMER | 0.015827916 | 0.005523337 | 0.026132495 | 2.81E-05 |
| CUPP-dbCAN_3:HMMER | -0.013614905 | -0.023919484 | -0.003310326 | 0.000872786 |
| dbCAN_3:eCAMI-dbCAN_3:DIAMOND | -0.031079997 | -0.041378321 | -0.020781672 | 0 |
| dbCAN_4-dbCAN_3:DIAMOND | -0.002522343 | -0.012826922 | 0.007782236 | 0.999854508 |
| dbCAN_4:HMMER-dbCAN_3:DIAMOND | -0.017415385 | -0.027719964 | -0.007110806 | 1.73E-06 |
| dbCAN_4:DIAMOND-dbCAN_3:DIAMOND | 0.000688386 | -0.009616193 | 0.010992965 | 1 |
| dbCAN_4:dbCAN-sub-dbCAN_3:DIAMOND | -0.003084958 | -0.013389537 | 0.007219621 | 0.998860226 |
| CUPP-dbCAN_3:DIAMOND | -0.032527779 | -0.042832358 | -0.0222232 | 0 |
| dbCAN_4-dbCAN_3:eCAMI | 0.028557654 | 0.01825933 | 0.038855978 | 0 |
| dbCAN_4:HMMER-dbCAN_3:eCAMI | 0.013664612 | 0.003366288 | 0.023962936 | 0.000803191 |
| dbCAN_4:DIAMOND-dbCAN_3:eCAMI | 0.031768383 | 0.021470059 | 0.042066707 | 0 |
| dbCAN_4:dbCAN-sub-dbCAN_3:eCAMI | 0.027995038 | 0.017696714 | 0.038293363 | 0 |
| CUPP-dbCAN_3:eCAMI | -0.001447782 | -0.011746106 | 0.008850542 | 0.999999692 |
| dbCAN_4:HMMER-dbCAN_4 | -0.014893042 | -0.025197621 | -0.004588463 | 0.000128016 |
| dbCAN_4:DIAMOND-dbCAN_4 | 0.003210729 | -0.00709385 | 0.013515308 | 0.998314999 |
| dbCAN_4:dbCAN-sub-dbCAN_4 | -0.000562615 | -0.010867194 | 0.009741963 | 1 |
| CUPP-dbCAN_4 | -0.030005436 | -0.040310015 | -0.019700857 | 0 |
| dbCAN_4:DIAMOND-dbCAN_4:HMMER | 0.018103771 | 0.007799192 | 0.02840835 | 4.75E-07 |
| dbCAN_4:dbCAN-sub-dbCAN_4:HMMER | 0.014330427 | 0.004025848 | 0.024635006 | 0.000304818 |
| CUPP-dbCAN_4:HMMER | -0.015112394 | -0.025416973 | -0.004807815 | 9.04E-05 |
| dbCAN_4:dbCAN-sub-dbCAN_4:DIAMOND | -0.003773344 | -0.014077923 | 0.006531235 | 0.992410048 |
| CUPP-dbCAN_4:DIAMOND | -0.033216165 | -0.043520744 | -0.022911586 | 0 |
| CUPP-dbCAN_4:dbCAN-sub | -0.029442821 | -0.039747399 | -0.019138242 | 0 |

## SI Table 21: Tukey HSD test to measure the statistically significant difference between the mean sensitivity for overall CAZy class classification (overleaf)

Tukey HSD test of the mean sensitivity between the classifiers, reporting the lower and upper 95% confidence interval, and the adjusted p-value (P-Adj).

| Classifier 1 - vs - Classifier 2 | Mean Difference | Lower 95% Confidence Interval | Upper 95% Confidence Interval | Adjusted P-value |
|---|---|---|---|---|
| dbCAN_2:HMMER-dbCAN_2 | -0.066067016 | -0.113803661 | -0.018330371 | 0.000338217 |
| dbCAN_2:DIAMOND-dbCAN_2 | -0.0001844 | -0.047921045 | 0.047552245 | 1 |
| dbCAN_2:Hotpep-dbCAN_2 | -0.071132015 | -0.11886866 | -0.02339537 | 6.12E-05 |
| dbCAN_3-dbCAN_2 | 0.045124247 | -0.002612399 | 0.092860892 | 0.085736247 |
| dbCAN_3:HMMER-dbCAN_2 | -0.048082121 | -0.095818766 | -0.000345476 | 0.046391849 |
| dbCAN_3:DIAMOND-dbCAN_2 | 0.076041479 | 0.028304834 | 0.123778124 | 1.03E-05 |
| dbCAN_3:eCAMI-dbCAN_2 | -0.062842561 | -0.110550231 | -0.015134891 | 0.000925733 |
| dbCAN_4-dbCAN_2 | 0.061812478 | 0.014075833 | 0.109549123 | 0.001281703 |
| dbCAN_4:HMMER-dbCAN_2 | -0.030633376 | -0.078370021 | 0.017103269 | 0.644902757 |
| dbCAN_4:DIAMOND-dbCAN_2 | 0.084488547 | 0.036751901 | 0.132225192 | 3.67E-07 |
| dbCAN_4:dbCAN-sub-dbCAN_2 | 0.063455161 | 0.015718515 | 0.111191806 | 0.000775159 |
| CUPP-dbCAN_2 | -0.159072605 | -0.20680925 | -0.11133596 | 0 |
| dbCAN_2:DIAMOND-dbCAN_2:HMMER | 0.065882616 | 0.018145971 | 0.113619261 | 0.000359037 |
| dbCAN_2:Hotpep-dbCAN_2:HMMER | -0.005064999 | -0.052801644 | 0.042671646 | 0.999999988 |
| dbCAN_3-dbCAN_2:HMMER | 0.111191262 | 0.063454617 | 0.158927907 | 0 |
| dbCAN_3:HMMER-dbCAN_2:HMMER | 0.017984894 | -0.029751751 | 0.06572154 | 0.990242322 |
| dbCAN_3:DIAMOND-dbCAN_2:HMMER | 0.142108494 | 0.094371849 | 0.18984514 | 0 |
| dbCAN_3:eCAMI-dbCAN_2:HMMER | 0.003224454 | -0.044483216 | 0.050932124 | 1 |
| dbCAN_4-dbCAN_2:HMMER | 0.127879494 | 0.080142849 | 0.175616139 | 0 |
| dbCAN_4:HMMER-dbCAN_2:HMMER | 0.035433639 | -0.012303006 | 0.083170285 | 0.401669508 |
| dbCAN_4:DIAMOND-dbCAN_2:HMMER | 0.150555562 | 0.102818917 | 0.198292207 | 0 |
| dbCAN_4:dbCAN-sub-dbCAN_2:HMMER | 0.129522176 | 0.081785531 | 0.177258821 | 0 |
| CUPP-dbCAN_2:HMMER | -0.093005589 | -0.140742235 | -0.045268944 | 9.04E-09 |
| dbCAN_2:Hotpep-dbCAN_2:DIAMOND | -0.070947615 | -0.11868426 | -0.02321097 | 6.53E-05 |
| dbCAN_3-dbCAN_2:DIAMOND | 0.045308646 | -0.002427999 | 0.093045292 | 0.082666368 |
| dbCAN_3:HMMER-dbCAN_2:DIAMOND | -0.047897721 | -0.095634366 | -0.000161076 | 0.048289009 |
| dbCAN_3:DIAMOND-dbCAN_2:DIAMOND | 0.076225879 | 0.028489234 | 0.123962524 | 9.63E-06 |
| dbCAN_3:eCAMI-dbCAN_2:DIAMOND | -0.062658161 | -0.110365831 | -0.014950491 | 0.000979681 |
| dbCAN_4-dbCAN_2:DIAMOND | 0.061996878 | 0.014260233 | 0.109733523 | 0.001212242 |
| dbCAN_4:HMMER-dbCAN_2:DIAMOND | -0.030448976 | -0.078185621 | 0.017287669 | 0.654138413 |
| dbCAN_4:DIAMOND-dbCAN_2:DIAMOND | 0.084672946 | 0.036936301 | 0.132409592 | 3.40E-07 |
| dbCAN_4:dbCAN-sub-dbCAN_2:DIAMOND | 0.063639561 | 0.015902915 | 0.111376206 | 0.000731945 |
| CUPP-dbCAN_2:DIAMOND | -0.158888205 | -0.20662485 | -0.11115156 | 0 |
| dbCAN_3-dbCAN_2:Hotpep | 0.116256261 | 0.068519616 | 0.163992907 | 0 |
| dbCAN_3:HMMER-dbCAN_2:Hotpep | 0.023049894 | -0.024686751 | 0.070786539 | 0.93024628 |
| dbCAN_3:DIAMOND-dbCAN_2:Hotpep | 0.147173494 | 0.099436849 | 0.194910139 | 0 |
| dbCAN_3:eCAMI-dbCAN_2:Hotpep | 0.008289454 | -0.039418216 | 0.055997124 | 0.999996567 |
| dbCAN_4-dbCAN_2:Hotpep | 0.132944493 | 0.085207848 | 0.180681138 | 0 |
| dbCAN_4:HMMER-dbCAN_2:Hotpep | 0.040498639 | -0.007238006 | 0.088235284 | 0.196822411 |
| dbCAN_4:DIAMOND-dbCAN_2:Hotpep | 0.155620561 | 0.107883916 | 0.203357207 | 0 |
| dbCAN_4:dbCAN-sub-dbCAN_2:Hotpep | 0.134587176 | 0.08685053 | 0.182323821 | 0 |
| CUPP-dbCAN_2:Hotpep | -0.08794059 | -0.135677235 | -0.040203945 | 8.53E-08 |
| dbCAN_3:HMMER-dbCAN_3 | -0.093206368 | -0.140943013 | -0.045469723 | 8.25E-09 |
| dbCAN_3:DIAMOND-dbCAN_3 | 0.030917232 | -0.016819413 | 0.078653877 | 0.630582857 |
| dbCAN_3:eCAMI-dbCAN_3 | -0.107966808 | -0.155674478 | -0.060259138 | 2.32E-12 |
| dbCAN_4-dbCAN_3 | 0.016688232 | -0.031048414 | 0.064424877 | 0.995007574 |
| dbCAN_4:HMMER-dbCAN_3 | -0.075757623 | -0.123494268 | -0.028020978 | 1.15E-05 |
| dbCAN_4:DIAMOND-dbCAN_3 | 0.0393643 | -0.008372345 | 0.087100945 | 0.235223773 |
| dbCAN_4:dbCAN-sub-dbCAN_3 | 0.018330914 | -0.029405731 | 0.066067559 | 0.988484945 |
| CUPP-dbCAN_3 | -0.204196852 | -0.251933497 | -0.156460206 | 0 |
| dbCAN_3:DIAMOND-dbCAN_3:HMMER | 0.1241236 | 0.076386955 | 0.171860245 | 0 |
| dbCAN_3:eCAMI-dbCAN_3:HMMER | -0.01476044 | -0.06246811 | 0.03294723 | 0.998426558 |
| dbCAN_4-dbCAN_3:HMMER | 0.109894599 | 0.062157954 | 0.157631244 | 0 |
| dbCAN_4:HMMER-dbCAN_3:HMMER | 0.017448745 | -0.0302879 | 0.06518539 | 0.992531177 |
| dbCAN_4:DIAMOND-dbCAN_3:HMMER | 0.132570668 | 0.084834023 | 0.180307313 | 0 |
| dbCAN_4:dbCAN-sub-dbCAN_3:HMMER | 0.111537282 | 0.063800637 | 0.159273927 | 0 |
| CUPP-dbCAN_3:HMMER | -0.110990484 | -0.158727129 | -0.063253839 | 0 |
| dbCAN_3:eCAMI-dbCAN_3:DIAMOND | -0.13888404 | -0.18659171 | -0.09117637 | 0 |
| dbCAN_4-dbCAN_3:DIAMOND | -0.014229001 | -0.061965646 | 0.033507644 | 0.998908325 |
| dbCAN_4:HMMER-dbCAN_3:DIAMOND | -0.106674855 | -0.1544115 | -0.05893821 | 8.26E-12 |
| dbCAN_4:DIAMOND-dbCAN_3:DIAMOND | 0.008447068 | -0.039289577 | 0.056183713 | 0.999995787 |
| dbCAN_4:dbCAN-sub-dbCAN_3:DIAMOND | -0.012586318 | -0.060322963 | 0.035150327 | 0.999684104 |
| CUPP-dbCAN_3:DIAMOND | -0.235114084 | -0.282850729 | -0.187377439 | 0 |
| dbCAN_4-dbCAN_3:eCAMI | 0.124655039 | 0.076947369 | 0.172362709 | 0 |
| dbCAN_4:HMMER-dbCAN_3:eCAMI | 0.032209185 | -0.015498485 | 0.079916855 | 0.563352879 |
| dbCAN_4:DIAMOND-dbCAN_3:eCAMI | 0.147331108 | 0.099623438 | 0.195038778 | 0 |
| dbCAN_4:dbCAN-sub-dbCAN_3:eCAMI | 0.126297722 | 0.078590052 | 0.174005392 | 0 |
| CUPP-dbCAN_3:eCAMI | -0.096230044 | -0.143937714 | -0.048522374 | 1.98E-09 |
| dbCAN_4:HMMER-dbCAN_4 | -0.092445854 | -0.140182499 | -0.044709209 | 1.17E-08 |
| dbCAN_4:DIAMOND-dbCAN_4 | 0.022676068 | -0.025060577 | 0.070412714 | 0.937866955 |
| dbCAN_4:dbCAN-sub-dbCAN_4 | 0.001642683 | -0.046093963 | 0.049379328 | 1 |
| CUPP-dbCAN_4 | -0.220885083 | -0.268621728 | -0.173148438 | 0 |
| dbCAN_4:DIAMOND-dbCAN_4:HMMER | 0.115121923 | 0.067385278 | 0.162858568 | 0 |
| dbCAN_4:dbCAN-sub-dbCAN_4:HMMER | 0.094088537 | 0.046351892 | 0.141825182 | 5.51E-09 |
| CUPP-dbCAN_4:HMMER | -0.128439229 | -0.176175874 | -0.080702584 | 0 |
| dbCAN_4:dbCAN-sub-dbCAN_4:DIAMOND | -0.021033386 | -0.068770031 | 0.026703259 | 0.964477906 |
| CUPP-dbCAN_4:DIAMOND | -0.243561152 | -0.291297797 | -0.195824506 | 0 |
| CUPP-dbCAN_4:dbCAN-sub | -0.222527766 | -0.270264411 | -0.174791121 | 0 |

## SI Table 22: Tukey HSD test to measure the statistically significant difference between the mean specificity for overall CAZy class classification (overleaf)

Tukey HSD test of the mean specificity between the classifiers, reporting the lower and upper 95% confidence interval, and the adjusted p-value (P-Adj).

| Classifier 1 - vs - Classifier 2 | Mean Difference | Lower 95% Confidence Interval | Upper 95% Confidence Interval | Adjusted P-value |
|---|---|---|---|---|
| dbCAN_2:HMMER-dbCAN_2 | 0.000548179 | -0.006408343 | 0.0075047 | 1 |
| dbCAN_2:DIAMOND-dbCAN_2 | -0.001306374 | -0.008262895 | 0.005650147 | 0.999991813 |
| dbCAN_2:Hotpep-dbCAN_2 | -0.020031326 | -0.026987847 | -0.013074805 | 0 |
| dbCAN_3-dbCAN_2 | 0.000110534 | -0.006845987 | 0.007067055 | 1 |
| dbCAN_3:HMMER-dbCAN_2 | 0.000800543 | -0.006155978 | 0.007757064 | 0.99999997 |
| dbCAN_3:DIAMOND-dbCAN_2 | -0.002473801 | -0.009430322 | 0.00448272 | 0.994167243 |
| dbCAN_3:eCAMI-dbCAN_2 | -0.010426366 | -0.017378665 | -0.003474067 | 5.21E-05 |
| dbCAN_4-dbCAN_2 | 0.000306274 | -0.006650247 | 0.007262795 | 1 |
| dbCAN_4:HMMER-dbCAN_2 | 0.000755899 | -0.006200622 | 0.007712421 | 0.999999985 |
| dbCAN_4:DIAMOND-dbCAN_2 | -0.001794266 | -0.008750787 | 0.005162255 | 0.999748441 |
| dbCAN_4:dbCAN-sub-dbCAN_2 | -0.000266477 | -0.007222998 | 0.006690044 | 1 |
| CUPP-dbCAN_2 | 0.001992667 | -0.004963854 | 0.008949188 | 0.999265487 |
| dbCAN_2:DIAMOND-dbCAN_2:HMMER | -0.001854552 | -0.008811073 | 0.005101969 | 0.999646007 |
| dbCAN_2:Hotpep-dbCAN_2:HMMER | -0.020579505 | -0.027536026 | -0.013622984 | 0 |
| dbCAN_3-dbCAN_2:HMMER | -0.000437645 | -0.007394166 | 0.006518876 | 1 |
| dbCAN_3:HMMER-dbCAN_2:HMMER | 0.000252365 | -0.006704157 | 0.007208886 | 1 |
| dbCAN_3:DIAMOND-dbCAN_2:HMMER | -0.003021979 | -0.009978501 | 0.003934542 | 0.968183952 |
| dbCAN_3:eCAMI-dbCAN_2:HMMER | -0.010974545 | -0.017926843 | -0.004022246 | 1.33E-05 |
| dbCAN_4-dbCAN_2:HMMER | -0.000241905 | -0.007198426 | 0.006714616 | 1 |
| dbCAN_4:HMMER-dbCAN_2:HMMER | 0.000207721 | -0.0067488 | 0.007164242 | 1 |
| dbCAN_4:DIAMOND-dbCAN_2:HMMER | -0.002342445 | -0.009298966 | 0.004614077 | 0.996470812 |
| dbCAN_4:dbCAN-sub-dbCAN_2:HMMER | -0.000814655 | -0.007771177 | 0.006141866 | 0.999999963 |
| CUPP-dbCAN_2:HMMER | 0.001444488 | -0.005512033 | 0.00840101 | 0.999975183 |
| dbCAN_2:Hotpep-dbCAN_2:DIAMOND | -0.018724953 | -0.025681474 | -0.011768431 | 0 |
| dbCAN_3-dbCAN_2:DIAMOND | 0.001416908 | -0.005539614 | 0.008373429 | 0.999979908 |
| dbCAN_3:HMMER-dbCAN_2:DIAMOND | 0.002106917 | -0.004849604 | 0.009063438 | 0.998722434 |
| dbCAN_3:DIAMOND-dbCAN_2:DIAMOND | -0.001167427 | -0.008123948 | 0.005789094 | 0.999997681 |
| dbCAN_3:eCAMI-dbCAN_2:DIAMOND | -0.009119992 | -0.016072291 | -0.002167694 | 0.001002602 |
| dbCAN_4-dbCAN_2:DIAMOND | 0.001612647 | -0.005343874 | 0.008569169 | 0.999918334 |
| dbCAN_4:HMMER-dbCAN_2:DIAMOND | 0.002062273 | -0.004894248 | 0.009018794 | 0.998965662 |
| dbCAN_4:DIAMOND-dbCAN_2:DIAMOND | -0.000487892 | -0.007444413 | 0.006468629 | 1 |
| dbCAN_4:dbCAN-sub-dbCAN_2:DIAMOND | 0.001039897 | -0.005916624 | 0.007996418 | 0.999999378 |
| CUPP-dbCAN_2:DIAMOND | 0.003299041 | -0.00365748 | 0.010255562 | 0.938599257 |
| dbCAN_3-dbCAN_2:Hotpep | 0.02014186 | 0.013185339 | 0.027098381 | 0 |
| dbCAN_3:HMMER-dbCAN_2:Hotpep | 0.02083187 | 0.013875348 | 0.027788391 | 0 |
| dbCAN_3:DIAMOND-dbCAN_2:Hotpep | 0.017557526 | 0.010601004 | 0.024514047 | 0 |
| dbCAN_3:eCAMI-dbCAN_2:Hotpep | 0.00960496 | 0.002652661 | 0.016557259 | 0.000351219 |
| dbCAN_4-dbCAN_2:Hotpep | 0.0203376 | 0.013381079 | 0.027294121 | 0 |
| dbCAN_4:HMMER-dbCAN_2:Hotpep | 0.020787226 | 0.013830704 | 0.027743747 | 0 |
| dbCAN_4:DIAMOND-dbCAN_2:Hotpep | 0.01823706 | 0.011280539 | 0.025193581 | 0 |
| dbCAN_4:dbCAN-sub-dbCAN_2:Hotpep | 0.019764849 | 0.012808328 | 0.026721371 | 0 |
| CUPP-dbCAN_2:Hotpep | 0.022023993 | 0.015067472 | 0.028980515 | 0 |
| dbCAN_3:HMMER-dbCAN_3 | 0.000690009 | -0.006266512 | 0.007646531 | 0.999999995 |
| dbCAN_3:DIAMOND-dbCAN_3 | -0.002584335 | -0.009540856 | 0.004372187 | 0.991375097 |
| dbCAN_3:eCAMI-dbCAN_3 | -0.0105369 | -0.017489199 | -0.003584601 | 3.98E-05 |
| dbCAN_4-dbCAN_3 | 0.00019574 | -0.006760781 | 0.007152261 | 1 |
| dbCAN_4:HMMER-dbCAN_3 | 0.000645366 | -0.006311156 | 0.007601887 | 0.999999998 |
| dbCAN_4:DIAMOND-dbCAN_3 | -0.0019048 | -0.008861321 | 0.005051721 | 0.999534543 |
| dbCAN_4:dbCAN-sub-dbCAN_3 | -0.000377011 | -0.007333532 | 0.00657951 | 1 |
| CUPP-dbCAN_3 | 0.001882133 | -0.005074388 | 0.008838654 | 0.999588119 |
| dbCAN_3:DIAMOND-dbCAN_3:HMMER | -0.003274344 | -0.010230865 | 0.003682177 | 0.941826459 |
| dbCAN_3:eCAMI-dbCAN_3:HMMER | -0.011226909 | -0.018179208 | -0.004274611 | 6.96E-06 |
| dbCAN_4-dbCAN_3:HMMER | -0.000494269 | -0.007450791 | 0.006462252 | 1 |
| dbCAN_4:HMMER-dbCAN_3:HMMER | -4.46E-05 | -0.007001165 | 0.006911877 | 1 |
| dbCAN_4:DIAMOND-dbCAN_3:HMMER | -0.002594809 | -0.00955133 | 0.004361712 | 0.991062009 |
| dbCAN_4:dbCAN-sub-dbCAN_3:HMMER | -0.00106702 | -0.008023541 | 0.005889501 | 0.999999165 |
| CUPP-dbCAN_3:HMMER | 0.001192124 | -0.005764397 | 0.008148645 | 0.999997063 |
| dbCAN_3:eCAMI-dbCAN_3:DIAMOND | -0.007952565 | -0.014904864 | -0.001000267 | 0.009729908 |
| dbCAN_4-dbCAN_3:DIAMOND | 0.002780075 | -0.004176447 | 0.009736596 | 0.983827291 |
| dbCAN_4:HMMER-dbCAN_3:DIAMOND | 0.0032297 | -0.003726821 | 0.010186221 | 0.947351793 |
| dbCAN_4:DIAMOND-dbCAN_3:DIAMOND | 0.000679535 | -0.006276986 | 0.007636056 | 0.999999996 |
| dbCAN_4:dbCAN-sub-dbCAN_3:DIAMOND | 0.002207324 | -0.004749197 | 0.009163845 | 0.997991264 |
| CUPP-dbCAN_3:DIAMOND | 0.004466468 | -0.002490053 | 0.011422989 | 0.644090436 |
| dbCAN_4-dbCAN_3:eCAMI | 0.01073264 | 0.003780341 | 0.017684939 | 2.45E-05 |
| dbCAN_4:HMMER-dbCAN_3:eCAMI | 0.011182266 | 0.004229967 | 0.018134564 | 7.81E-06 |
| dbCAN_4:DIAMOND-dbCAN_3:eCAMI | 0.0086321 | 0.001679801 | 0.015584399 | 0.002710172 |
| dbCAN_4:dbCAN-sub-dbCAN_3:eCAMI | 0.010159889 | 0.003207591 | 0.017112188 | 9.85E-05 |
| CUPP-dbCAN_3:eCAMI | 0.012419033 | 0.005466734 | 0.019371332 | 2.65E-07 |
| dbCAN_4:HMMER-dbCAN_4 | 0.000449626 | -0.006506896 | 0.007406147 | 1 |
| dbCAN_4:DIAMOND-dbCAN_4 | -0.00210054 | -0.009057061 | 0.004855981 | 0.99875992 |
| dbCAN_4:dbCAN-sub-dbCAN_4 | -0.000572751 | -0.007529272 | 0.006383771 | 0.999999999 |
| CUPP-dbCAN_4 | 0.001686393 | -0.005270128 | 0.008642914 | 0.999868649 |
| dbCAN_4:DIAMOND-dbCAN_4:HMMER | -0.002550165 | -0.009506687 | 0.004406356 | 0.992334971 |
| dbCAN_4:dbCAN-sub-dbCAN_4:HMMER | -0.001022376 | -0.007978897 | 0.005934145 | 0.999999488 |
| CUPP-dbCAN_4:HMMER | 0.001236768 | -0.005719753 | 0.008193289 | 0.999995559 |
| dbCAN_4:dbCAN-sub-dbCAN_4:DIAMOND | 0.001527789 | -0.005428732 | 0.00848431 | 0.99995433 |
| CUPP-dbCAN_4:DIAMOND | 0.003786933 | -0.003169588 | 0.010743454 | 0.848028607 |
| CUPP-dbCAN_4:dbCAN-sub | 0.002259144 | -0.004697377 | 0.009215665 | 0.997491831 |

## SI Table 23: Tukey HSD test to measure the statistically significant difference between the mean precision for overall CAZy class classification (overleaf)

Tukey HSD test of the mean precision between the classifiers, reporting the lower and upper 95% confidence interval, and the adjusted p-value (P-Adj).

| Classifier 1 - vs - Classifier 2 | Mean Difference | Lower 95% Confidence Interval | Upper 95% Confidence Interval | Adjusted P-value |
|---|---|---|---|---|
| dbCAN_2:HMMER-dbCAN_2 | -0.002329371 | -0.04392239 | 0.039263648 | 1 |
| dbCAN_2:DIAMOND-dbCAN_2 | -0.008126356 | -0.049719375 | 0.033466663 | 0.999987297 |
| dbCAN_2:Hotpep-dbCAN_2 | -0.100833952 | -0.142426971 | -0.059240933 | 0 |
| dbCAN_3-dbCAN_2 | 0.00856838 | -0.033024639 | 0.050161399 | 0.999977246 |
| dbCAN_3:HMMER-dbCAN_2 | 0.00047366 | -0.041119359 | 0.042066678 | 1 |
| dbCAN_3:DIAMOND-dbCAN_2 | 0.00163763 | -0.039955389 | 0.043230649 | 1 |
| dbCAN_3:eCAMI-dbCAN_2 | -0.068790874 | -0.110358647 | -0.027223101 | 3.34E-06 |
| dbCAN_4-dbCAN_2 | 0.011654395 | -0.029938624 | 0.053247414 | 0.999411765 |
| dbCAN_4:HMMER-dbCAN_2 | 0.004782388 | -0.036810631 | 0.046375407 | 0.99999997 |
| dbCAN_4:DIAMOND-dbCAN_2 | 0.006525129 | -0.03506789 | 0.048118148 | 0.99999892 |
| dbCAN_4:dbCAN-sub-dbCAN_2 | 0.007537174 | -0.034055845 | 0.049130193 | 0.9999945 |
| CUPP-dbCAN_2 | -0.178257037 | -0.219850056 | -0.136664018 | 0 |
| dbCAN_2:DIAMOND-dbCAN_2:HMMER | -0.005796985 | -0.047390004 | 0.035796034 | 0.999999721 |
| dbCAN_2:Hotpep-dbCAN_2:HMMER | -0.098504581 | -0.1400976 | -0.056911562 | 0 |
| dbCAN_3-dbCAN_2:HMMER | 0.010897751 | -0.030695268 | 0.05249077 | 0.999703967 |
| dbCAN_3:HMMER-dbCAN_2:HMMER | 0.002803031 | -0.038789988 | 0.044396049 | 1 |
| dbCAN_3:DIAMOND-dbCAN_2:HMMER | 0.003967001 | -0.037626018 | 0.04556002 | 0.999999997 |
| dbCAN_3:eCAMI-dbCAN_2:HMMER | -0.066461503 | -0.108029276 | -0.02489373 | 9.28E-06 |
| dbCAN_4-dbCAN_2:HMMER | 0.013983766 | -0.027609253 | 0.055576785 | 0.996521612 |
| dbCAN_4:HMMER-dbCAN_2:HMMER | 0.007111759 | -0.03448126 | 0.048704778 | 0.999997136 |
| dbCAN_4:DIAMOND-dbCAN_2:HMMER | 0.0088545 | -0.032738519 | 0.050447519 | 0.999967428 |
| dbCAN_4:dbCAN-sub-dbCAN_2:HMMER | 0.009866545 | -0.031726474 | 0.051459564 | 0.999895629 |
| CUPP-dbCAN_2:HMMER | -0.175927666 | -0.217520685 | -0.134334647 | 0 |
| dbCAN_2:Hotpep-dbCAN_2:DIAMOND | -0.092707596 | -0.134300615 | -0.051114577 | 1.00E-11 |
| dbCAN_3-dbCAN_2:DIAMOND | 0.016694736 | -0.024898283 | 0.058287755 | 0.983225593 |
| dbCAN_3:HMMER-dbCAN_2:DIAMOND | 0.008600015 | -0.032993003 | 0.050193034 | 0.999976309 |
| dbCAN_3:DIAMOND-dbCAN_2:DIAMOND | 0.009763986 | -0.031829033 | 0.051357004 | 0.999906611 |
| dbCAN_3:eCAMI-dbCAN_2:DIAMOND | -0.060664518 | -0.102232291 | -0.019096746 | 0.000101695 |
| dbCAN_4-dbCAN_2:DIAMOND | 0.019780751 | -0.021812268 | 0.06137377 | 0.937347655 |
| dbCAN_4:HMMER-dbCAN_2:DIAMOND | 0.012908744 | -0.028684275 | 0.054501763 | 0.998378187 |
| dbCAN_4:DIAMOND-dbCAN_2:DIAMOND | 0.014651485 | -0.026941534 | 0.056244504 | 0.994648318 |
| dbCAN_4:dbCAN-sub-dbCAN_2:DIAMOND | 0.01566353 | -0.025929489 | 0.057256549 | 0.990278966 |
| CUPP-dbCAN_2:DIAMOND | -0.170130681 | -0.2117237 | -0.128537662 | 0 |
| dbCAN_3-dbCAN_2:Hotpep | 0.109402332 | 0.067809313 | 0.150995351 | 0 |
| dbCAN_3:HMMER-dbCAN_2:Hotpep | 0.101307612 | 0.059714593 | 0.142900631 | 0 |
| dbCAN_3:DIAMOND-dbCAN_2:Hotpep | 0.102471582 | 0.060878563 | 0.144064601 | 0 |
| dbCAN_3:eCAMI-dbCAN_2:Hotpep | 0.032043078 | -0.009524695 | 0.073610851 | 0.338533229 |
| dbCAN_4-dbCAN_2:Hotpep | 0.112488347 | 0.070895328 | 0.154081366 | 0 |
| dbCAN_4:HMMER-dbCAN_2:Hotpep | 0.10561634 | 0.064023321 | 0.147209359 | 0 |
| dbCAN_4:DIAMOND-dbCAN_2:Hotpep | 0.107359081 | 0.065766062 | 0.1489521 | 0 |
| dbCAN_4:dbCAN-sub-dbCAN_2:Hotpep | 0.108371126 | 0.066778107 | 0.149964145 | 0 |
| CUPP-dbCAN_2:Hotpep | -0.077423085 | -0.119016104 | -0.035830066 | 5.73E-08 |
| dbCAN_3:HMMER-dbCAN_3 | -0.00809472 | -0.049687739 | 0.033498298 | 0.999987834 |
| dbCAN_3:DIAMOND-dbCAN_3 | -0.00693075 | -0.048523769 | 0.034662269 | 0.999997859 |
| dbCAN_3:eCAMI-dbCAN_3 | -0.077359254 | -0.118927027 | -0.035791482 | 5.77E-08 |
| dbCAN_4-dbCAN_3 | 0.003086015 | -0.038507004 | 0.044679034 | 1 |
| dbCAN_4:HMMER-dbCAN_3 | -0.003785992 | -0.045379011 | 0.037807027 | 0.999999998 |
| dbCAN_4:DIAMOND-dbCAN_3 | -0.002043251 | -0.04363627 | 0.039549768 | 1 |
| dbCAN_4:dbCAN-sub-dbCAN_3 | -0.001031206 | -0.042624225 | 0.040561813 | 1 |
| CUPP-dbCAN_3 | -0.186825417 | -0.228418436 | -0.145232398 | 0 |
| dbCAN_3:DIAMOND-dbCAN_3:HMMER | 0.00116397 | -0.040429049 | 0.042756989 | 1 |
| dbCAN_3:eCAMI-dbCAN_3:HMMER | -0.069264534 | -0.110832306 | -0.027696761 | 2.70E-06 |
| dbCAN_4-dbCAN_3:HMMER | 0.011180735 | -0.030412283 | 0.052773754 | 0.999614524 |
| dbCAN_4:HMMER-dbCAN_3:HMMER | 0.004308729 | -0.03728429 | 0.045901747 | 0.999999991 |
| dbCAN_4:DIAMOND-dbCAN_3:HMMER | 0.006051469 | -0.035541549 | 0.047644488 | 0.999999544 |
| dbCAN_4:dbCAN-sub-dbCAN_3:HMMER | 0.007063514 | -0.034529504 | 0.048656533 | 0.999997347 |
| CUPP-dbCAN_3:HMMER | -0.178730696 | -0.220323715 | -0.137137678 | 0 |
| dbCAN_3:eCAMI-dbCAN_3:DIAMOND | -0.070428504 | -0.111996277 | -0.028860731 | 1.59E-06 |
| dbCAN_4-dbCAN_3:DIAMOND | 0.010016765 | -0.031576254 | 0.051609784 | 0.999877486 |
| dbCAN_4:HMMER-dbCAN_3:DIAMOND | 0.003144758 | -0.038448261 | 0.044737777 | 1 |
| dbCAN_4:DIAMOND-dbCAN_3:DIAMOND | 0.004887499 | -0.03670552 | 0.046480518 | 0.999999961 |
| dbCAN_4:dbCAN-sub-dbCAN_3:DIAMOND | 0.005899544 | -0.035693475 | 0.047492563 | 0.999999659 |
| CUPP-dbCAN_3:DIAMOND | -0.179894667 | -0.221487686 | -0.138301648 | 0 |
| dbCAN_4-dbCAN_3:eCAMI | 0.080445269 | 0.038877497 | 0.122013042 | 1.20E-08 |
| dbCAN_4:HMMER-dbCAN_3:eCAMI | 0.073573262 | 0.03200549 | 0.115141035 | 3.67E-07 |
| dbCAN_4:DIAMOND-dbCAN_3:eCAMI | 0.075316003 | 0.033748231 | 0.116883776 | 1.58E-07 |
| dbCAN_4:dbCAN-sub-dbCAN_3:eCAMI | 0.076328048 | 0.034760275 | 0.117895821 | 9.64E-08 |
| CUPP-dbCAN_3:eCAMI | -0.109466163 | -0.151033935 | -0.06789839 | 0 |
| dbCAN_4:HMMER-dbCAN_4 | -0.006872007 | -0.048465026 | 0.034721012 | 0.999998056 |
| dbCAN_4:DIAMOND-dbCAN_4 | -0.005129266 | -0.046722285 | 0.036463753 | 0.999999932 |
| dbCAN_4:dbCAN-sub-dbCAN_4 | -0.004117221 | -0.04571024 | 0.037475798 | 0.999999995 |
| CUPP-dbCAN_4 | -0.189911432 | -0.231504451 | -0.148318413 | 0 |
| dbCAN_4:DIAMOND-dbCAN_4:HMMER | 0.001742741 | -0.039850278 | 0.04333576 | 1 |
| dbCAN_4:dbCAN-sub-dbCAN_4:HMMER | 0.002754786 | -0.038838233 | 0.044347805 | 1 |
| CUPP-dbCAN_4:HMMER | -0.183039425 | -0.224632444 | -0.141446406 | 0 |
| dbCAN_4:dbCAN-sub-dbCAN_4:DIAMOND | 0.001012045 | -0.040580974 | 0.042605064 | 1 |
| CUPP-dbCAN_4:DIAMOND | -0.184782166 | -0.226375185 | -0.143189147 | 0 |
| CUPP-dbCAN_4:dbCAN-sub | -0.185794211 | -0.22738723 | -0.144201192 | 0 |

# 5 Comparison of general trends in performance between the CAZy classes

The data is collated by statistical parameter, for example, the specificity scores for all CAZy classes are plotted in a single figure and are grouped by CAZy class (e.g. SI figure 18). This section facilitates comparing the values of the same statistical parameter across multiple CAZy classes.

In the online the report the data is grouped in this manner, as well as collated by CAZy class, then grouped (or facet wrapped) by statistical parameter (e.g. SI figure **??**). Therefore, a single plot is generated for CAzy class, plotting all five of the statistical parameters (specificity, sensitivity, precision, F1-score, and accuracy). This section facilitates the evaluating the performance of the classifiers for a specific class.

## SI Table 25: Output of two-way ANOVAs testing for statistically significant differences between the classifiers and the CAZy classes.

Reporting the statistical parameter the two-way ANOVA was completed for; the source of variation (term); degrees of freedom (Df); Sum of squares (SumSeq); Mean square (MeanSq); F-statistic; and the corresponding p-value.

| Statistical Parameter | Term | Df | SumSeq | MeanSq | F-Statistic | P-value |
|---|---|---|---|---|---|---|
| **Specificity** | CAZy_class | 5 | 0.107652 | 0.02153 | 27.4802 | 1.48E-27 |
| | Prediction_tool | 12 | 0.184722 | 0.015393 | 19.64731 | 1.32E-42 |
| | CAZy_class:Prediction_tool | 60 | 0.592856 | 0.009881 | 12.61143 | 5.00E-112 |
| | Residuals | 5266 | 4.125857 | 0.000783 | NA | NA |
| **Sensitivity** | CAZy_class | 5 | 63.65174428 | 12.73034886 | 581.8355523 | 0 |
| | Prediction_tool | 12 | 26.25175868 | 2.187646556 | 99.98551938 | 4.55E-224 |
| | CAZy_class:Prediction_tool | 60 | 48.39891499 | 0.806648583 | 36.867554 | 0 |
| | Residuals | 5266 | 115.218152 | 0.021879634 | NA | NA |
| **Precision** | CAZy_class | 5 | 24.94465662 | 4.988931325 | 299.6566887 | 5.14E-283 |
| | Prediction_tool | 12 | 16.17508995 | 1.347924162 | 80.96212691 | 2.40E-183 |
| | CAZy_class:Prediction_tool | 60 | 59.91939131 | 0.998656522 | 59.9836091 | 0 |
| | Residuals | 5266 | 87.67270463 | 0.016648824 | NA | NA |
| **F1-score** | CAZy_class | 5 | 52.16660846 | 10.43332169 | 611.627543 | 0 |
| | Prediction_tool | 12 | 20.4549855 | 1.704582125 | 99.92688883 | 6.04E-224 |
| | CAZy_class:Prediction_tool | 60 | 45.32440215 | 0.755406702 | 44.2838397 | 0 |
| | Residuals | 5266 | 89.82896971 | 0.017058293 | NA | NA |
| **Accuracy** | CAZy_class | 5 | 1.630458 | 0.326092 | 211.3496 | 5.21E-206 |
| | Prediction_tool | 12 | 1.07743 | 0.089786 | 58.19286 | 9.60E-133 |
| | CAZy_class:Prediction_tool | 60 | 0.834128 | 0.013902 | 9.01039 | 5.19E-74 |
| | Residuals | 5266 | 8.124917 | 0.001543 | NA | NA |

## 5.1 F1-score

## SI Figure 14: F1-score of binary CAZy class classification performance per CAZy class

F1-score was calculated for each test set per CAZy class. The mean F1-score was then calculated across all test sets and 95% confidence interval (CI) was calculated for each CAZy class, and is plotted in SI figure 14.

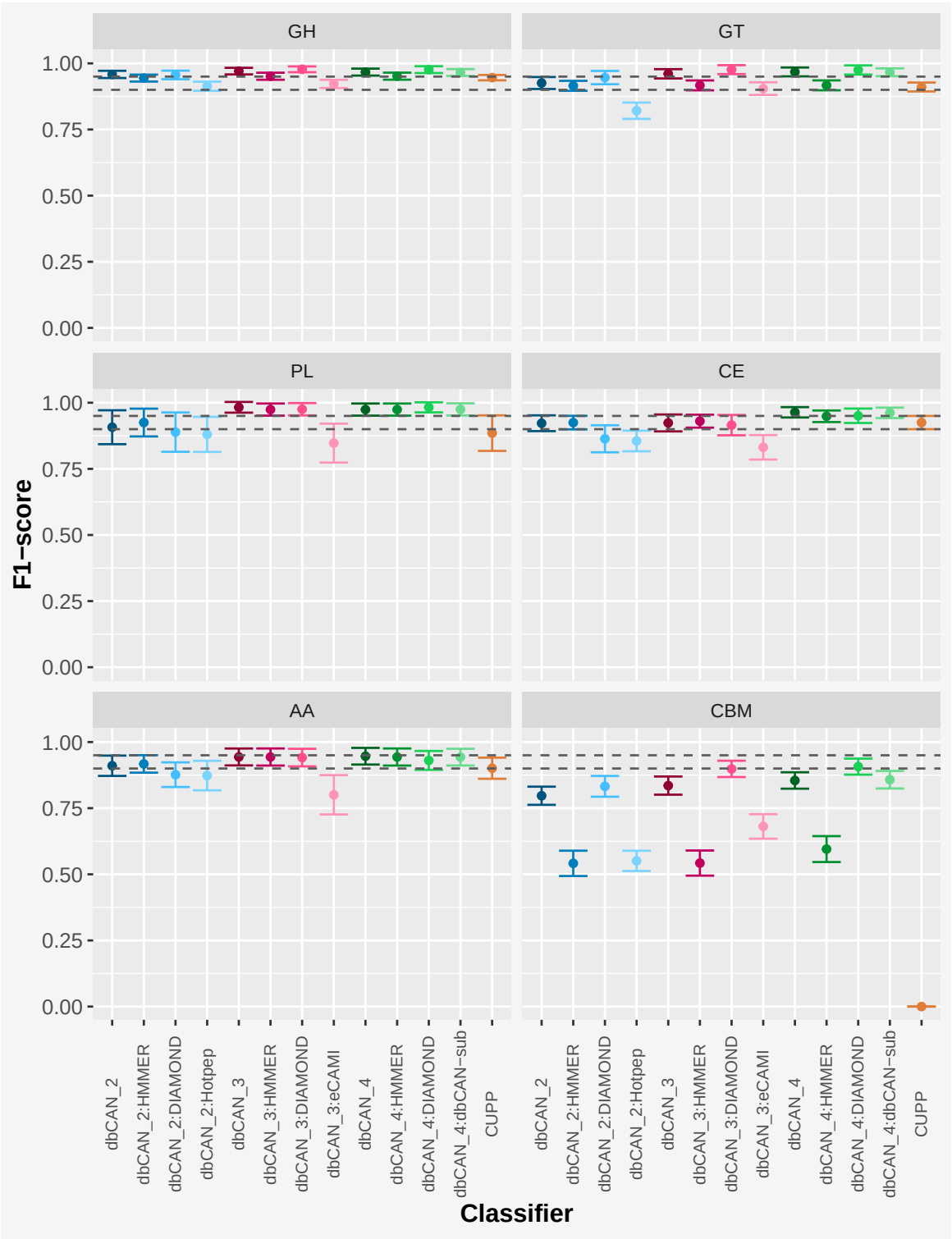



Figure 14: The mean F1-score and 95% confidence interval (CI) of binary CAZy class classification per CAZy class, across all test sets. Where a tool is incorporated into a CAZyme classifier, the tool name is prefixed with the respective CAZyme classifier. dbCAN_2 and associated tools are coloured blue; dbCAN_3 pink; dbCAN_4 green; and CUPP orange.

## SI Figure 15: F1-score of binary CAZy class classification performance per test set

F1-score was calculated for each test set per CAZy class, and was plotted in SI figure 15.

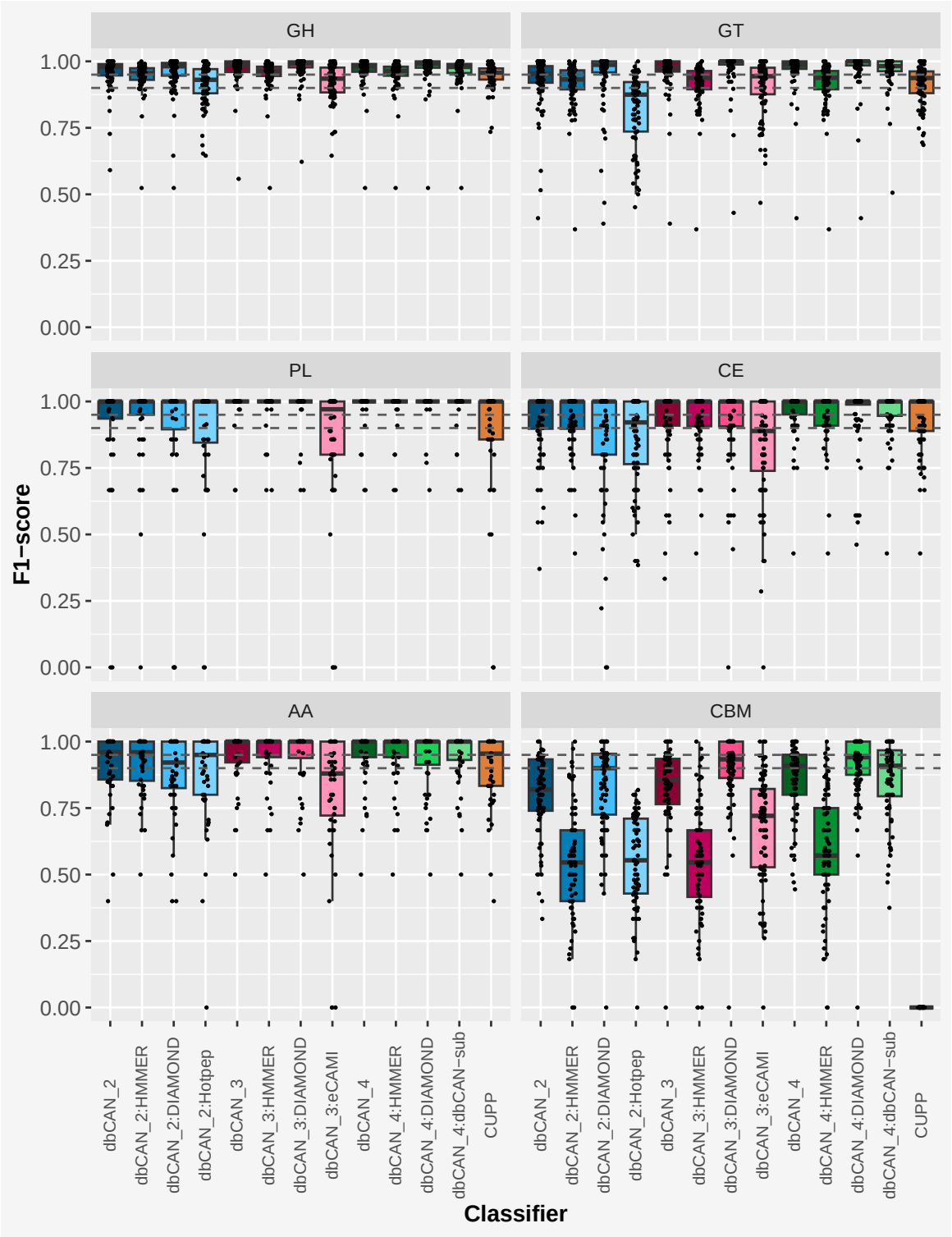



Figure 15: The F1-score of the binary classification of each CAZy class was calculated per test set, and is plotted as a one-dimensional scatter plot overlaying a box and whisker plot. Where a tool is incorporated into a CAZyme classifier, the tool name is prefixed with the respective CAZyme classifier. dbCAN_2 and associated tools are coloured blue; dbCAN_3 pink; dbCAN_4 green; and CUPP orange.

## 5.2 The distribution of F1-scores for CAZy class classification

### SI Table 26: The distribution of F1-scores for CAZy class classification (overleaf)

For instances where an F1-score was achieved for a test set for a given CAZy class, the percentage of the instances with an F1-score greater than a given threshold per CAZy class and across all CAZy classes (i.e. regardless of the CAZy class) are listed in SI table 6. The mean and standard deviation across all classifiers is also reported.

| Prediction Tool | GH (%) | | | GT (%) | | | PL (%) | | | CE (%) | | | AA (%) | | | CBM (%) | | | All CAZy Classes (%) | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | 1 | ≥0.9 | ≤0.75 | 1 | ≥0.9 | ≤0.75 | 1 | ≥0.9 | ≤0.75 | 1 | ≥0.9 | ≤0.75 | 1 | ≥0.9 | ≤0.75 | 1 | ≥0.9 | ≤0.75 | 1 | ≥0.9 | ≤0.75 |
| dbCAN_2 | 20.000 | 91.250 | 8.750 | 12.500 | 81.250 | 18.750 | 68.085 | 76.596 | 23.404 | 59.211 | 75.000 | 26.316 | 47.917 | 64.583 | 35.417 | 5.000 | 33.750 | 67.500 | 31.630 | 70.316 | 30.170 |
| dbCAN_2:HMMER | 1.250 | 88.750 | 11.250 | 2.500 | 72.500 | 28.750 | 70.213 | 78.723 | 21.277 | 51.316 | 75.000 | 26.316 | 47.917 | 70.833 | 31.250 | 1.250 | 6.250 | 93.750 | 24.088 | 63.747 | 36.983 |
| dbCAN_2:DIAMOND | 25.000 | 88.750 | 11.250 | 45.000 | 85.000 | 15.000 | 65.957 | 74.468 | 25.532 | 52.632 | 64.474 | 36.842 | 45.833 | 52.083 | 47.917 | 18.750 | 51.250 | 51.250 | 39.903 | 70.316 | 30.414 |
| dbCAN_2:Hotpep | 5.000 | 67.500 | 32.500 | 1.250 | 40.000 | 62.500 | 61.702 | 65.957 | 34.043 | 42.105 | 52.632 | 47.368 | 45.833 | 58.333 | 41.667 | 0.000 | 0.000 | 100.000 | 21.411 | 45.012 | 55.474 |
| dbCAN_3 | 31.250 | 96.250 | 3.750 | 31.250 | 91.250 | 8.750 | 89.362 | 95.745 | 4.255 | 60.526 | 77.632 | 23.684 | 68.750 | 79.167 | 20.833 | 12.500 | 42.500 | 58.750 | 44.039 | 79.319 | 21.168 |
| dbCAN_3:HMMER | 5.000 | 91.250 | 8.750 | 2.500 | 73.750 | 26.250 | 85.106 | 91.489 | 8.511 | 52.632 | 76.316 | 25.000 | 66.667 | 81.250 | 18.750 | 1.250 | 6.250 | 93.750 | 28.954 | 67.397 | 32.847 |
| dbCAN_3:DIAMOND | 43.750 | 95.000 | 5.000 | 68.750 | 93.750 | 6.250 | 87.234 | 91.489 | 8.511 | 63.158 | 77.632 | 25.000 | 70.833 | 75.000 | 25.000 | 28.750 | 67.500 | 36.250 | 57.421 | 83.455 | 17.762 |
| dbCAN_3:eCAMI | 8.750 | 67.500 | 32.500 | 11.250 | 70.000 | 31.250 | 50.000 | 54.167 | 45.833 | 39.474 | 48.684 | 51.316 | 29.167 | 45.833 | 56.250 | 5.000 | 13.750 | 86.250 | 21.359 | 50.000 | 50.485 |
| dbCAN_4 | 18.750 | 95.000 | 5.000 | 37.500 | 92.500 | 7.500 | 85.106 | 91.489 | 8.511 | 72.368 | 89.474 | 10.526 | 68.750 | 83.333 | 16.667 | 11.250 | 52.500 | 48.750 | 44.282 | 83.455 | 16.788 |
| dbCAN_4:HMMER | 5.000 | 91.250 | 8.750 | 2.500 | 73.750 | 26.250 | 85.106 | 91.489 | 8.511 | 60.526 | 84.211 | 15.789 | 66.667 | 81.250 | 18.750 | 1.250 | 8.750 | 92.500 | 30.414 | 69.343 | 30.900 |
| dbCAN_4:DIAMOND | 40.000 | 95.000 | 5.000 | 66.250 | 95.000 | 5.000 | 89.362 | 93.617 | 6.383 | 75.000 | 88.158 | 13.158 | 66.667 | 75.000 | 25.000 | 33.750 | 70.000 | 33.750 | 59.124 | 86.375 | 14.599 |
| dbCAN_4:dbCAN-sub | 16.250 | 96.250 | 3.750 | 32.500 | 92.500 | 7.500 | 87.234 | 91.489 | 8.511 | 71.053 | 88.158 | 11.842 | 66.667 | 79.167 | 20.833 | 23.750 | 51.250 | 48.750 | 45.012 | 82.725 | 17.275 |
| CUPP | 3.750 | 90.000 | 10.000 | 8.750 | 67.500 | 35.000 | 59.574 | 68.085 | 31.915 | 55.263 | 69.737 | 31.579 | 47.917 | 62.500 | 37.500 | 0.000 | 0.000 | 100.000 | 25.061 | 58.637 | 42.092 |
| Mean | 17.212 | 88.750 | 11.250 | 24.808 | 79.135 | 21.442 | 75.695 | 81.908 | 18.092 | 58.097 | 74.393 | 26.518 | 56.891 | 69.872 | 30.449 | 10.962 | 31.058 | 70.096 | 36.361 | 70.007 | 30.535 |
| Standard Deviation | 13.721 | 9.418 | 9.418 | 23.137 | 14.834 | 15.604 | 13.022 | 12.699 | 12.699 | 10.423 | 12.350 | 12.277 | 12.748 | 11.689 | 12.025 | 11.289 | 25.138 | 24.081 | 12.435 | 12.646 | 12.620 |

## SI Table 27: Output of Tukey HSD test for statistically significant differences between the mean f1-score of CAZy class and classifiers, reporting only tests with a p-value <0.05 (Overleaf).

Specifically, only hits where a statistically significant difference (p-value<0.05) is found between the same tool (i.e. Classifier 1 and Classifier 2 are the same) for different CAZy classes (i.e. Class 1 and Class 2 are different). The complete dataset is available in the online repository. Here the pair of CAZy classes, pair of classifiers/tools, the mean difference (Mean Diff), lower and upper 95% confidence interval and the adjusted p-value (Adj.p-value) are reported.

| Class 1 | Class 2 | Classifier 1 | Classifier 2 | MeanDiff | Lower | Upper | Adj.p-value |
|---|---|---|---|---|---|---|---|
| CBM | AA | dbCAN_2:HMMER | dbCAN_2:HMMER | -0.375905907 | -0.475977101 | -0.275834714 | 0 |
| CE | CBM | dbCAN_2:HMMER | dbCAN_2:HMMER | 0.383466736 | 0.295669626 | 0.471263845 | 0 |
| GH | CBM | dbCAN_2:HMMER | dbCAN_2:HMMER | 0.402998053 | 0.316333857 | 0.489662249 | 0 |
| GT | CBM | dbCAN_2:HMMER | dbCAN_2:HMMER | 0.37380939 | 0.287145194 | 0.460473585 | 0 |
| PL | CBM | dbCAN_2:HMMER | dbCAN_2:HMMER | 0.383660164 | 0.282925801 | 0.484394527 | 0 |
| CBM | AA | dbCAN_3:eCAMI | dbCAN_3:eCAMI | -0.119624969 | -0.219696163 | -0.019553776 | 0.001405052 |
| GH | AA | dbCAN_3:eCAMI | dbCAN_3:eCAMI | 0.121759445 | 0.021688251 | 0.221830638 | 0.000898346 |
| GT | AA | dbCAN_3:eCAMI | dbCAN_3:eCAMI | 0.103982884 | 0.003911691 | 0.204054078 | 0.026607441 |
| CE | CBM | dbCAN_3:eCAMI | dbCAN_3:eCAMI | 0.150399028 | 0.062601919 | 0.238196137 | 2.23E-09 |
| GH | CBM | dbCAN_3:eCAMI | dbCAN_3:eCAMI | 0.241384414 | 0.154720219 | 0.32804861 | 0 |
| GT | CBM | dbCAN_3:eCAMI | dbCAN_3:eCAMI | 0.223607854 | 0.136943658 | 0.310272049 | 0 |
| PL | CBM | dbCAN_3:eCAMI | dbCAN_3:eCAMI | 0.166301316 | 0.066230122 | 0.266372509 | 1.03E-08 |
| GH | CE | dbCAN_3:eCAMI | dbCAN_3:eCAMI | 0.090985386 | 0.003188277 | 0.178782495 | 0.027868528 |
| CBM | AA | dbCAN_3:HMMER | dbCAN_3:HMMER | -0.40096441 | -0.501035603 | -0.300893216 | 0 |
| CE | CBM | dbCAN_3:HMMER | dbCAN_3:HMMER | 0.387745566 | 0.299948457 | 0.475542675 | 0 |
| GH | CBM | dbCAN_3:HMMER | dbCAN_3:HMMER | 0.408957593 | 0.322293397 | 0.495621788 | 0 |
| GT | CBM | dbCAN_3:HMMER | dbCAN_3:HMMER | 0.374381299 | 0.287717103 | 0.461045495 | 0 |
| PL | CBM | dbCAN_3:HMMER | dbCAN_3:HMMER | 0.431612246 | 0.330877883 | 0.532346608 | 0 |
| CBM | AA | dbCAN_2 | dbCAN_2 | -0.11337027 | -0.213441463 | -0.013299077 | 0.004904921 |
| CE | CBM | dbCAN_2 | dbCAN_2 | 0.125363435 | 0.037566326 | 0.213160544 | 6.45E-06 |
| GH | CBM | dbCAN_2 | dbCAN_2 | 0.161245812 | 0.074581617 | 0.247910008 | 2.04E-11 |
| GT | CBM | dbCAN_2 | dbCAN_2 | 0.128785434 | 0.042121238 | 0.215449629 | 1.43E-06 |
| PL | CBM | dbCAN_2 | dbCAN_2 | 0.110270174 | 0.009535811 | 0.211004537 | 0.010052331 |
| GH | CBM | dbCAN_2:DIAMOND | dbCAN_2:DIAMOND | 0.123743506 | 0.03707931 | 0.210407701 | 6.46E-06 |
| GT | CBM | dbCAN_2:DIAMOND | dbCAN_2:DIAMOND | 0.113758996 | 0.0270948 | 0.200423191 | 0.000106547 |
| GH | CE | dbCAN_2:DIAMOND | dbCAN_2:DIAMOND | 0.09262993 | 0.004832821 | 0.180427039 | 0.02029894 |
| CE | CBM | dbCAN_4 | dbCAN_4 | 0.109065404 | 0.021268295 | 0.196862513 | 0.000519226 |
| GH | CBM | dbCAN_4 | dbCAN_4 | 0.112331242 | 0.025667047 | 0.198995438 | 0.000155763 |
| GT | CBM | dbCAN_4 | dbCAN_4 | 0.113027431 | 0.026363235 | 0.199691626 | 0.000129521 |
| PL | CBM | dbCAN_4 | dbCAN_4 | 0.119361563 | 0.0186272 | 0.220095926 | 0.001743983 |
| CBM | AA | CUPP | CUPP | -0.901051314 | -1.001122507 | -0.80098012 | 0 |
| CE | CBM | CUPP | CUPP | 0.925011345 | 0.837214236 | 1.012808454 | 0 |
| GH | CBM | CUPP | CUPP | 0.946062049 | 0.859397853 | 1.032726244 | 0 |
| GT | CBM | CUPP | CUPP | 0.910715975 | 0.824051779 | 0.997380171 | 0 |
| PL | CBM | CUPP | CUPP | 0.885017551 | 0.784283188 | 0.985751914 | 0 |
| CE | CBM | dbCAN_4:sub | dbCAN_4:sub | 0.10408055 | 0.016283441 | 0.191877659 | 0.001724339 |
| GH | CBM | dbCAN_4:sub | dbCAN_4:sub | 0.107923683 | 0.021259487 | 0.194587878 | 0.000485674 |
| GT | CBM | dbCAN_4:sub | dbCAN_4:sub | 0.10875238 | 0.022088185 | 0.195416576 | 0.000393803 |
| PL | CBM | dbCAN_4:sub | dbCAN_4:sub | 0.11710031 | 0.016365948 | 0.217834673 | 0.002746433 |
| CBM | AA | dbCAN_3 | dbCAN_3 | -0.108217768 | -0.208288961 | -0.008146574 | 0.012777641 |
| CE | CBM | dbCAN_3 | dbCAN_3 | 0.088184286 | 0.000387177 | 0.175981395 | 0.046675598 |
| GH | CBM | dbCAN_3 | dbCAN_3 | 0.135356682 | 0.048692486 | 0.222020878 | 1.82E-07 |
| GT | CBM | dbCAN_3 | dbCAN_3 | 0.125197457 | 0.038533261 | 0.211861652 | 4.20E-06 |
| PL | CBM | dbCAN_3 | dbCAN_3 | 0.147195095 | 0.046460732 | 0.247929457 | 2.74E-06 |
| CBM | AA | dbCAN_2:Hotpep | dbCAN_2:Hotpep | -0.322460821 | -0.422532014 | -0.222389627 | 0 |
| CE | CBM | dbCAN_2:Hotpep | dbCAN_2:Hotpep | 0.304636124 | 0.216839015 | 0.392433233 | 0 |
| GH | CBM | dbCAN_2:Hotpep | dbCAN_2:Hotpep | 0.362805674 | 0.276141478 | 0.44946987 | 0 |
| GT | CBM | dbCAN_2:Hotpep | dbCAN_2:Hotpep | 0.27007451 | 0.183410314 | 0.356738705 | 0 |
| PL | CBM | dbCAN_2:Hotpep | dbCAN_2:Hotpep | 0.329437829 | 0.228703466 | 0.430172192 | 0 |
| GT | GH | dbCAN_2:Hotpep | dbCAN_2:Hotpep | -0.092731164 | -0.17939536 | -0.006066968 | 0.015646695 |
| CBM | AA | dbCAN_4:HMMER | dbCAN_4:HMMER | -0.348056264 | -0.448127457 | -0.24798507 | 0 |
| CE | CBM | dbCAN_4:HMMER | dbCAN_4:HMMER | 0.353329233 | 0.265532124 | 0.441126342 | 0 |
| GH | CBM | dbCAN_4:HMMER | dbCAN_4:HMMER | 0.356253601 | 0.269589405 | 0.442917796 | 0 |
| GT | CBM | dbCAN_4:HMMER | dbCAN_4:HMMER | 0.321670242 | 0.235006047 | 0.408334438 | 0 |
| PL | CBM | dbCAN_4:HMMER | dbCAN_4:HMMER | 0.378704099 | 0.277969736 | 0.479438462 | 0 |

## 5.3 Accuracy

### SI Figure 16: Accuracy of binary CAZy class classification performance per CAZy class

Accuracy was calculated for each test set per CAZy class. The mean accuracy was then calculated across all test sets and 95% confidence interval (CI) was calculated for each CAZy class, and is plotted in SI figure 16.

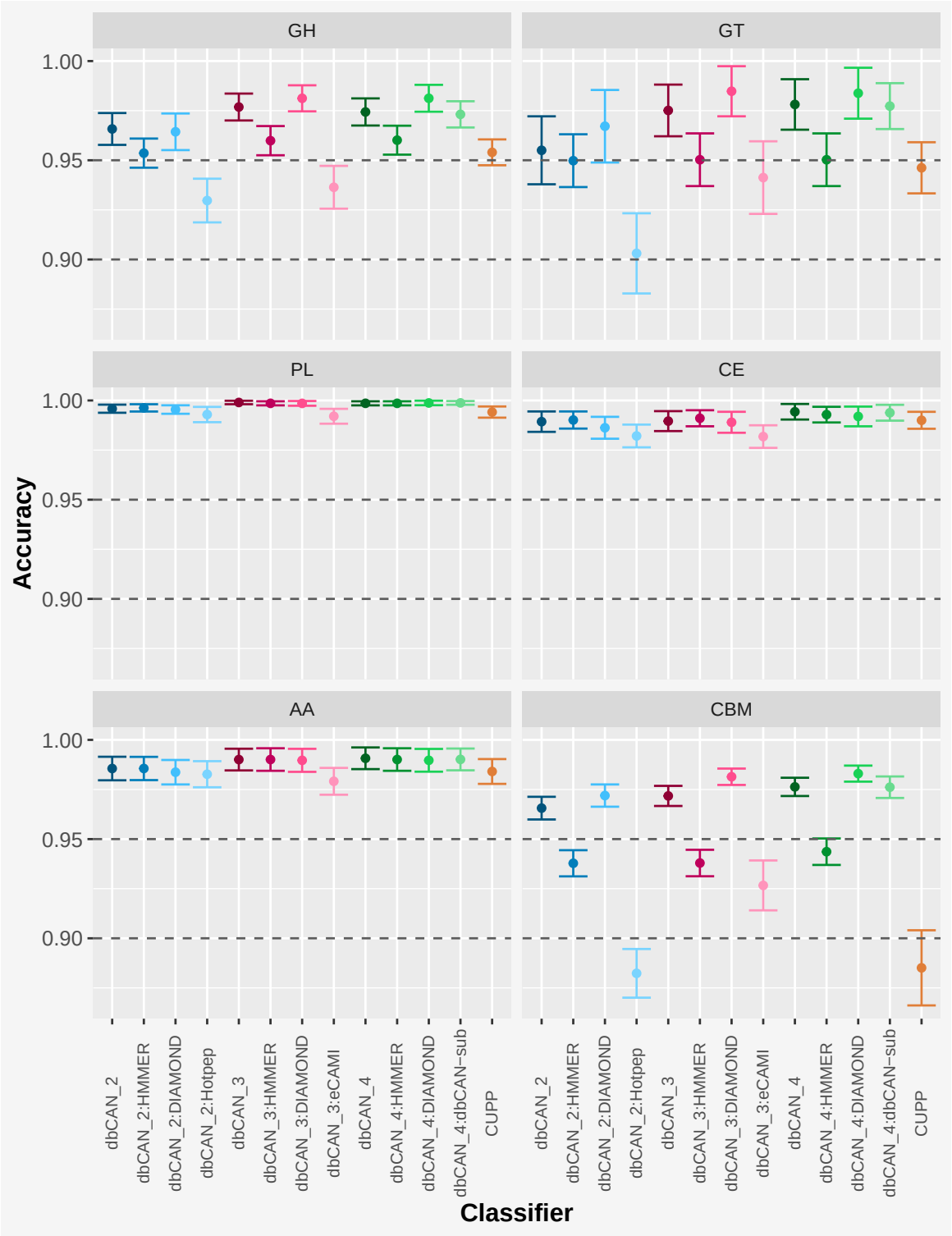



Figure 16: The mean accuracy and 95% confidence interval (CI) of binary CAZy class classification per CAZy class, across all test sets. Where a tool is incorporated into a CAZyme classifier, the tool name is prefixed with the respective CAZyme classifier. dbCAN_2 and associated tools are coloured blue; dbCAN_3 pink; dbCAN_4 green; and CUPP orange.

## SI Figure 17: Accuracy of binary CAZy class classification performance per test set

Accuracy was calculated for each test set per CAZy class, and was plotted in SI figure 17.


Figure 17: The accuracy of the binary classification of each CAZy class was calculated per test set, and is plotted as a one-dimensional scatter plot overlaying a box and whisker plot. Where a tool is incorporated into a CAZyme classifier, the tool name is prefixed with the respective CAZyme classifier. dbCAN_2 and associated tools are coloured blue; dbCAN_3 pink; dbCAN_4 green; and CUPP orange.

## SI Table 27: Output of Tukey HSD test for statistically significant differences between the mean accuracy of CAZy class and classifiers, reporting only tests with a p-value <0.05 (Overleaf).

Specifically, only hits where a statistically significant difference (p-value<0.05) is found between the same tool (i.e. Classifier 1 and Classifier 2 are the same) for different CAZy classes (i.e. Class 1 and Class 2 are different). The complete dataset is available in the online repository. Here the pair of CAZy classes, pair of classifiers/tools, the mean difference (Mean Diff), lower and upper 95% confidence interval and the adjusted p-value (Adj.p-value) are reported.

| Class 1 | Class 2 | Classifier 1 | Classifier 2 | MeanDiff | Lower | Upper | Adj.p-value |
|---|---|---|---|---|---|---|---|
| CBM | AA | dbCAN_2:HMMER | dbCAN_2:HMMER | -0.047778944 | -0.077875052 | -0.017682837 | 8.86E-08 |
| GH | AA | dbCAN_2:HMMER | dbCAN_2:HMMER | -0.031990494 | -0.062086601 | -0.001894387 | 0.017705966 |
| GT | AA | dbCAN_2:HMMER | dbCAN_2:HMMER | -0.035793717 | -0.065889824 | -0.00569761 | 0.001593277 |
| CE | CBM | dbCAN_2:HMMER | dbCAN_2:HMMER | 0.052342212 | 0.025937498 | 0.078746925 | 7.19E-14 |
| PL | CBM | dbCAN_2:HMMER | dbCAN_2:HMMER | 0.058461591 | 0.028166038 | 0.088757145 | 1.93E-12 |
| GH | CE | dbCAN_2:HMMER | dbCAN_2:HMMER | -0.036553761 | -0.062958475 | -0.010149047 | 1.92E-05 |
| GT | CE | dbCAN_2:HMMER | dbCAN_2:HMMER | -0.040356984 | -0.066761698 | -0.013952271 | 4.57E-07 |
| PL | GH | dbCAN_2:HMMER | dbCAN_2:HMMER | 0.042673141 | 0.012377587 | 0.072968694 | 1.05E-05 |
| PL | GT | dbCAN_2:HMMER | dbCAN_2:HMMER | 0.046476364 | 0.016180811 | 0.076771917 | 3.91E-07 |
| CBM | AA | dbCAN_3:eCAMI | dbCAN_3:eCAMI | -0.052458242 | -0.082554349 | -0.022362135 | 8.88E-10 |
| GH | AA | dbCAN_3:eCAMI | dbCAN_3:eCAMI | -0.04273419 | -0.072830298 | -0.012638083 | 7.89E-06 |
| GT | AA | dbCAN_3:eCAMI | dbCAN_3:eCAMI | -0.037855467 | -0.067951574 | -0.00775936 | 0.000369003 |
| CE | CBM | dbCAN_3:eCAMI | dbCAN_3:eCAMI | 0.055140654 | 0.02873594 | 0.081545368 | 0 |
| PL | CBM | dbCAN_3:eCAMI | dbCAN_3:eCAMI | 0.06537773 | 0.035281622 | 0.095473837 | 0 |
| GH | CE | dbCAN_3:eCAMI | dbCAN_3:eCAMI | -0.045416603 | -0.071821316 | -0.019011889 | 1.80E-09 |
| GT | CE | dbCAN_3:eCAMI | dbCAN_3:eCAMI | -0.040537879 | -0.066942592 | -0.014133165 | 3.79E-07 |
| PL | GH | dbCAN_3:eCAMI | dbCAN_3:eCAMI | 0.055653678 | 0.025557571 | 0.085749785 | 2.97E-11 |
| PL | GT | dbCAN_3:eCAMI | dbCAN_3:eCAMI | 0.050774954 | 0.020678847 | 0.080871061 | 4.88E-09 |
| CBM | AA | dbCAN_3:HMMER | dbCAN_3:HMMER | -0.052167679 | -0.082263786 | -0.022071572 | 1.20E-09 |
| GH | AA | dbCAN_3:HMMER | dbCAN_3:HMMER | -0.030223911 | -0.060320018 | -0.000127804 | 0.046795242 |
| GT | AA | dbCAN_3:HMMER | dbCAN_3:HMMER | -0.039833844 | -0.069929951 | -0.009737737 | 8.26E-05 |
| CE | CBM | dbCAN_3:HMMER | dbCAN_3:HMMER | 0.053102944 | 0.02669823 | 0.079507657 | 0 |
| PL | CBM | dbCAN_3:HMMER | dbCAN_3:HMMER | 0.060627181 | 0.030331627 | 0.090922734 | 0 |
| GH | CE | dbCAN_3:HMMER | dbCAN_3:HMMER | -0.031159176 | -0.057563889 | -0.004754462 | 0.001925802 |
| GT | CE | dbCAN_3:HMMER | dbCAN_3:HMMER | -0.040769109 | -0.067173823 | -0.014364395 | 2.98E-07 |
| PL | GH | dbCAN_3:HMMER | dbCAN_3:HMMER | 0.038683412 | 0.008387859 | 0.068978966 | 0.000241335 |
| PL | GT | dbCAN_3:HMMER | dbCAN_3:HMMER | 0.048293346 | 0.017997792 | 0.078588899 | 7.36E-08 |
| GT | AA | dbCAN_2 | dbCAN_2 | -0.030545917 | -0.060642024 | -0.00044981 | 0.039502578 |
| GT | CE | dbCAN_2 | dbCAN_2 | -0.034304686 | -0.060709399 | -0.007899972 | 0.000145344 |
| PL | GT | dbCAN_2 | dbCAN_2 | 0.040834608 | 0.010539055 | 0.071130161 | 4.64E-05 |
| PL | GH | dbCAN_2:DIAMOND | dbCAN_2:DIAMOND | 0.031074824 | 0.000779271 | 0.061370378 | 0.033186695 |
| CBM | AA | CUPP | CUPP | -0.098968757 | -0.129064864 | -0.06887265 | 0 |
| GT | AA | CUPP | CUPP | -0.037877063 | -0.06797317 | -0.007780956 | 0.000363197 |
| CE | CBM | CUPP | CUPP | 0.104901645 | 0.078496931 | 0.131306359 | 0 |
| GH | CBM | CUPP | CUPP | 0.068884464 | 0.042820471 | 0.094948457 | 0 |
| GT | CBM | CUPP | CUPP | 0.061091694 | 0.035027701 | 0.087155687 | 0 |
| PL | CBM | CUPP | CUPP | 0.109039838 | 0.078744285 | 0.139335392 | 0 |
| GH | CE | CUPP | CUPP | -0.036017181 | -0.062421895 | -0.009612468 | 3.15E-05 |
| GT | CE | CUPP | CUPP | -0.043809951 | -0.070214665 | -0.017405237 | 1.12E-08 |
| PL | GH | CUPP | CUPP | 0.040155374 | 0.009859821 | 0.070450928 | 7.90E-05 |
| PL | GT | CUPP | CUPP | 0.047948144 | 0.017652591 | 0.078243698 | 1.02E-07 |
| CBM | AA | dbCAN_2:Hotpep | dbCAN_2:Hotpep | -0.100356257 | -0.130452365 | -0.07026015 | 0 |
| GH | AA | dbCAN_2:Hotpep | dbCAN_2:Hotpep | -0.052991664 | -0.083087772 | -0.022895557 | 5.12E-10 |
| GT | AA | dbCAN_2:Hotpep | dbCAN_2:Hotpep | -0.079612082 | -0.109708189 | -0.049515975 | 0 |
| CE | CBM | dbCAN_2:Hotpep | dbCAN_2:Hotpep | 0.099783141 | 0.073378428 | 0.126187855 | 0 |
| GH | CBM | dbCAN_2:Hotpep | dbCAN_2:Hotpep | 0.047364593 | 0.0213006 | 0.073428586 | 8.48E-11 |
| PL | CBM | dbCAN_2:Hotpep | dbCAN_2:Hotpep | 0.110579312 | 0.080283758 | 0.140874865 | 0 |
| GH | CE | dbCAN_2:Hotpep | dbCAN_2:Hotpep | -0.052418548 | -0.078823262 | -0.026013835 | 3.32E-14 |
| GT | CE | dbCAN_2:Hotpep | dbCAN_2:Hotpep | -0.079038966 | -0.105443679 | -0.052634252 | 0 |
| GT | GH | dbCAN_2:Hotpep | dbCAN_2:Hotpep | -0.026620418 | -0.052684411 | -0.000556424 | 0.035633515 |
| PL | GH | dbCAN_2:Hotpep | dbCAN_2:Hotpep | 0.063214719 | 0.032919165 | 0.093510272 | 0 |
| PL | GT | dbCAN_2:Hotpep | dbCAN_2:Hotpep | 0.089835136 | 0.059539583 | 0.12013069 | 0 |
| CBM | AA | dbCAN_4:HMMER | dbCAN_4:HMMER | -0.046461002 | -0.076557109 | -0.016364895 | 3.00E-07 |
| GT | AA | dbCAN_4:HMMER | dbCAN_4:HMMER | -0.039837414 | -0.069933521 | -0.009741307 | 8.24E-05 |
| CE | CBM | dbCAN_4:HMMER | dbCAN_4:HMMER | 0.049228102 | 0.022823388 | 0.075632816 | 1.80E-11 |
| PL | CBM | dbCAN_4:HMMER | dbCAN_4:HMMER | 0.054920504 | 0.02462495 | 0.085216057 | 9.78E-11 |
| GH | CE | dbCAN_4:HMMER | dbCAN_4:HMMER | -0.032741011 | -0.059145724 | -0.006336297 | 0.000545552 |
| GT | CE | dbCAN_4:HMMER | dbCAN_4:HMMER | -0.042604514 | -0.069009228 | -0.0161998 | 4.23E-08 |
| PL | GH | dbCAN_4:HMMER | dbCAN_4:HMMER | 0.038433412 | 0.008137859 | 0.068728966 | 0.000290344 |
| PL | GT | dbCAN_4:HMMER | dbCAN_4:HMMER | 0.048296916 | 0.018001362 | 0.078592469 | 7.34E-08 |

## 5.4 Specificity

### SI Figure 18: Specificity of binary CAZy class classification performance per CAZy class

Specificity was calculated for each test set per CAZy class. The mean specificity was then calculated across all test sets and 95% confidence interval (CI) was calculated for each CAZy class, and is plotted in SI figure 18.

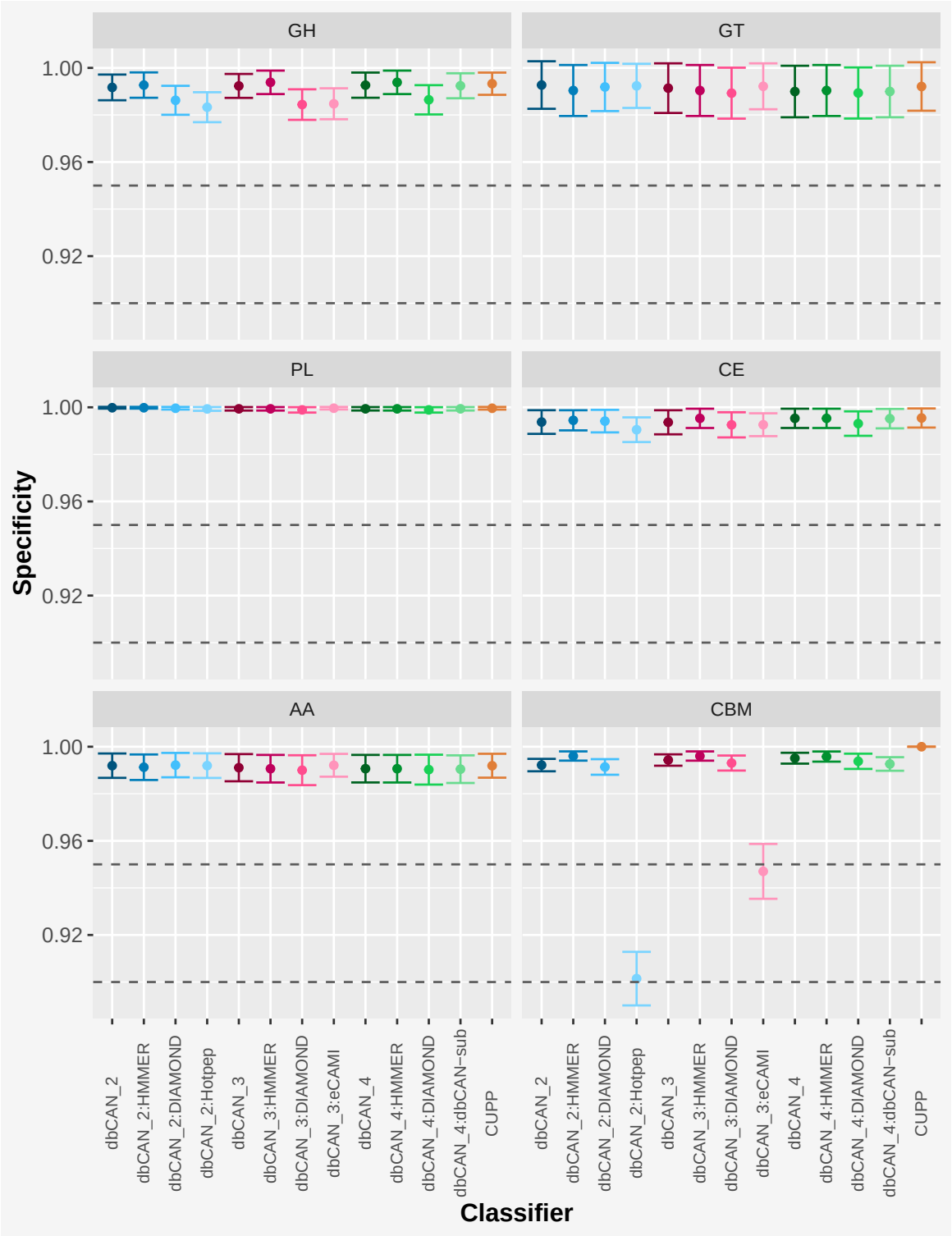



Figure 18: The mean specificity and 95% confidence interval (CI) of binary CAZy class classification per CAZy class, across all test sets. Where a tool is incorporated into a CAZyme classifier, the tool name is prefixed with the respective CAZyme classifier. dbCAN_2 and associated tools are coloured blue; dbCAN_3 pink; dbCAN_4 green; and CUPP orange.

## SI Figure 19: Specificity of binary CAZy class classification performance per test set

Specificity was calculated for each test set per CAZy class, and was plotted in SI figure 19.

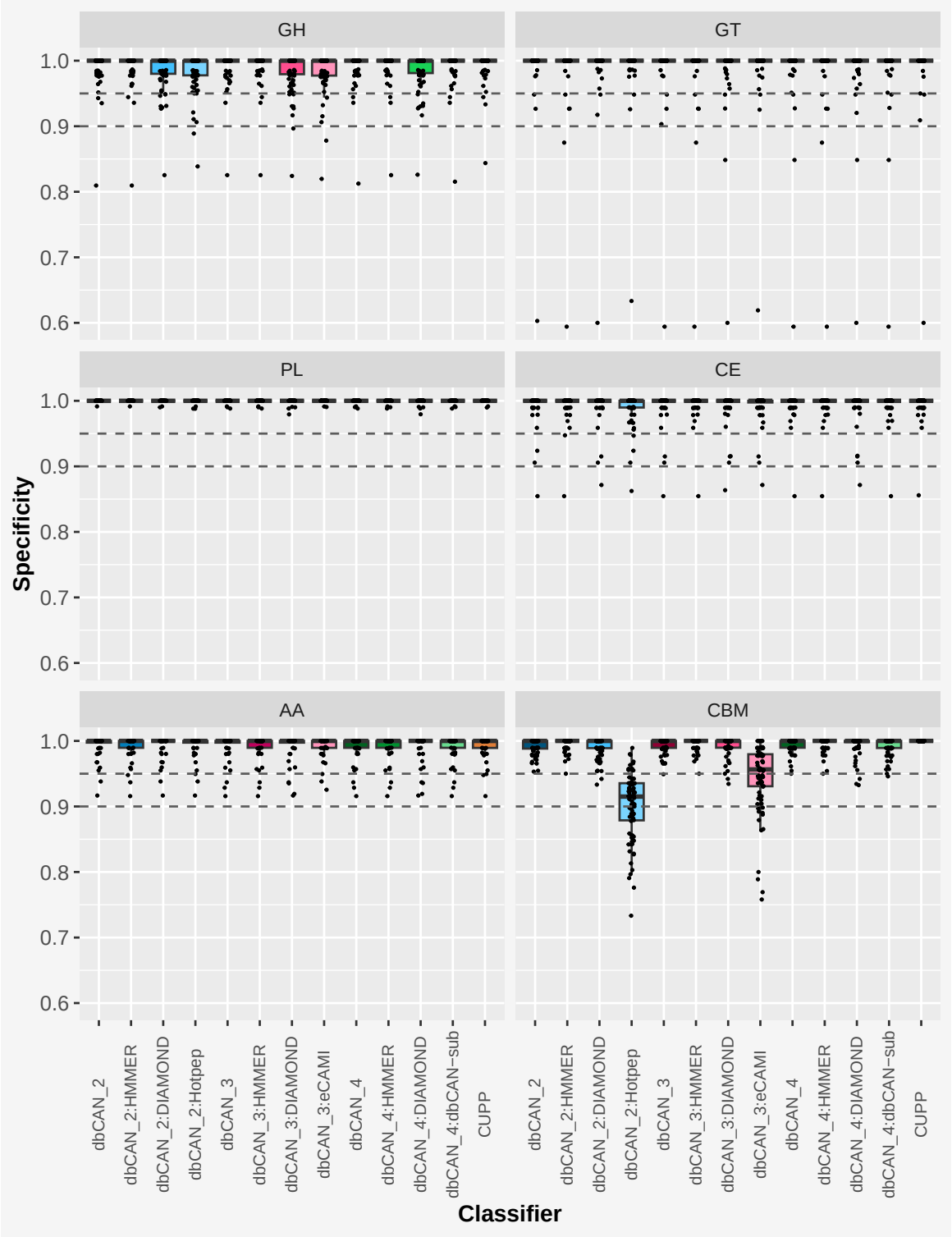



Figure 19: The specificity of the binary classification of each CAZy class was calculated per test set, and is plotted as a one-dimensional scatter plot overlaying a box and whisker plot. Where a tool is incorporated into a CAZyme classifier, the tool name is prefixed with the respective CAZyme classifier. dbCAN_2 and associated tools are coloured blue; dbCAN_3 pink; dbCAN_4 green; and CUPP orange.

## SI Table 28: Output of Tukey HSD test for statistically significant differences between the mean specificity of CAZy class and classifiers, reporting only tests with a p-value <0.05 (Overleaf).

Specifically, only hits where a statistically significant difference (p-value<0.05) is found between the same tool (i.e. Classifier 1 and Classifier 2 are the same) for different CAZy classes (i.e. Class 1 and Class 2 are different). The complete dataset is available in the online repository. Here the pair of CAZy classes, pair of classifiers/tools, the mean difference (Mean Diff), lower and upper 95% confidence interval and the adjusted p-value (Adj.p-value) are reported.

| Class 1 | Class 2 | Classifier 1 | Classifier 2 | MeanDiff | Lower | Upper | Adj.p-value |
|---|---|---|---|---|---|---|---|
| CBM | AA | dbCAN_3:eCAMI | dbCAN_3:eCAMI | -0.04508 | -0.06653 | -0.02364 | 0 |
| CE | CBM | dbCAN_3:eCAMI | dbCAN_3:eCAMI | 0.045572 | 0.026756 | 0.064388 | 0 |
| GH | CBM | dbCAN_3:eCAMI | dbCAN_3:eCAMI | 0.037729 | 0.019156 | 0.056303 | 0 |
| GT | CBM | dbCAN_3:eCAMI | dbCAN_3:eCAMI | 0.045148 | 0.026575 | 0.063721 | 0 |
| PL | CBM | dbCAN_3:eCAMI | dbCAN_3:eCAMI | 0.052597 | 0.03115 | 0.074043 | 0 |
| CBM | AA | dbCAN_2:Hotpep | dbCAN_2:Hotpep | -0.09052 | -0.11196 | -0.06907 | 0 |
| CE | CBM | dbCAN_2:Hotpep | dbCAN_2:Hotpep | 0.08903 | 0.070214 | 0.107846 | 0 |
| GH | CBM | dbCAN_2:Hotpep | dbCAN_2:Hotpep | 0.081846 | 0.063273 | 0.10042 | 0 |
| GT | CBM | dbCAN_2:Hotpep | dbCAN_2:Hotpep | 0.090919 | 0.072345 | 0.109492 | 0 |
| PL | CBM | dbCAN_2:Hotpep | dbCAN_2:Hotpep | 0.097861 | 0.076272 | 0.119449 | 0 |

## 5.5 Precision

### SI Figure 20: Precision of binary CAZy class classification performance per CAZy class

Precision was calculated for each test set per CAZy class. The mean precision was then calculated across all test sets and 95% confidence interval (CI) was calculated for each CAZy class, and is plotted in SI figure 20.

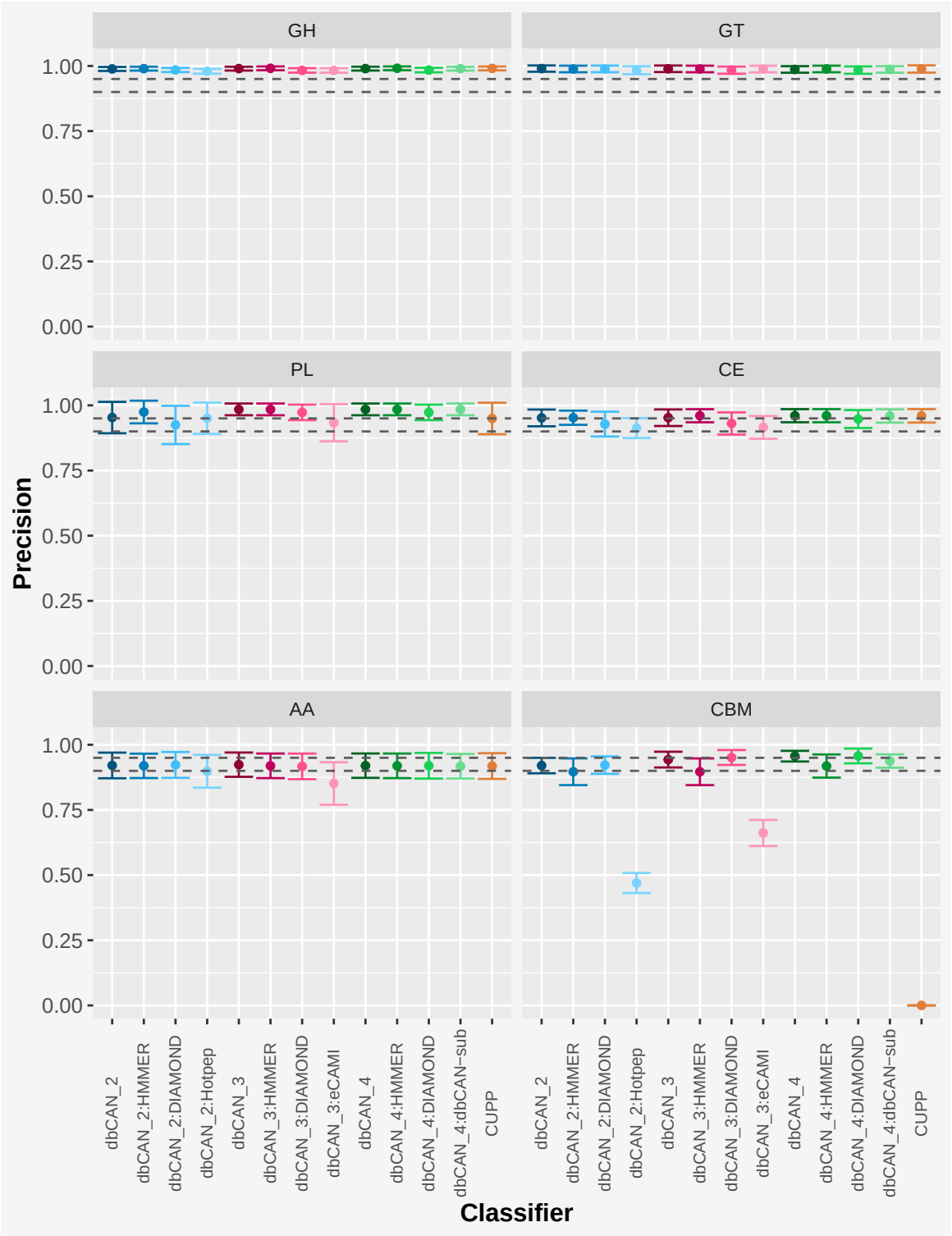



Figure 20: The mean precision and 95% confidence interval (CI) of binary CAZy class classification per CAZy class, across all test sets. Where a tool is incorporated into a CAZyme classifier, the tool name is prefixed with the respective CAZyme classifier. dbCAN_2 and associated tools are coloured blue; dbCAN_3 pink; dbCAN_4 green; and CUPP orange.

## SI Figure 21: Precision of binary CAZy class classification performance per test set

Precision was calculated for each test set per CAZy class, and was plotted in SI figure 21.

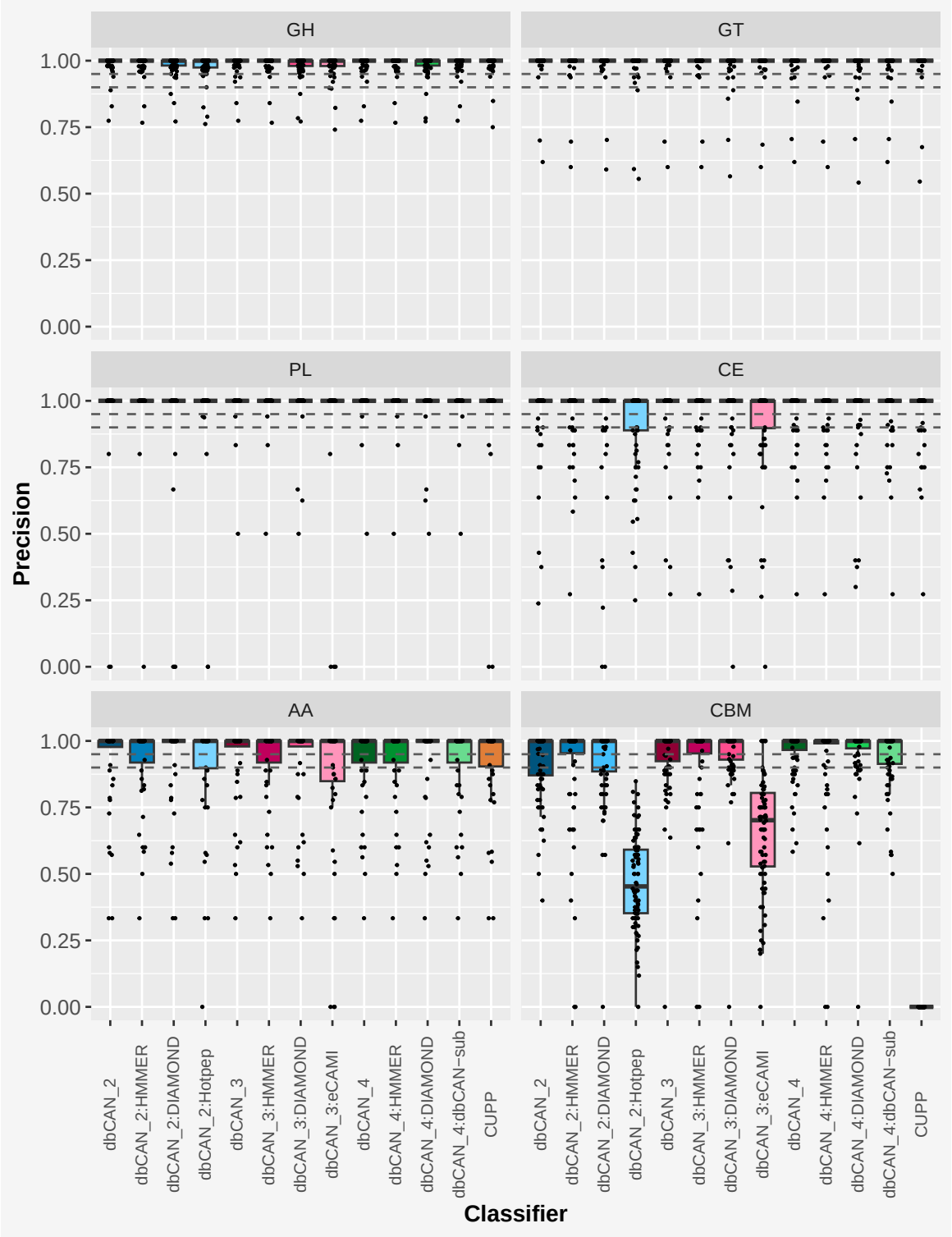



Figure 21: The precision of the binary classification of each CAZy class was calculated per test set, and is plotted as a one-dimensional scatter plot overlaying a box and whisker plot. Where a tool is incorporated into a CAZyme classifier, the tool name is prefixed with the respective CAZyme classifier. dbCAN_2 and associated tools are coloured blue; dbCAN_3 pink; dbCAN_4 green; and CUPP orange.

## SI Table 29: Output of Tukey HSD test for statistically significant differences between the mean precision of CAZy class and classifiers, reporting only tests with a p-value <0.05 (Overleaf).

Specifically, only hits where a statistically significant difference (p-value<0.05) is found between the same tool (i.e. Classifier 1 and Classifier 2 are the same) for different CAZy classes (i.e. Class 1 and Class 2 are different). The complete dataset is available in the online repository. Here the pair of CAZy classes, pair of classifiers/tools, the mean difference (Mean Diff), lower and upper 95% confidence interval and the adjusted p-value (Adj.p-value) are reported.

| Class 1 | Class 2 | Classifier 1 | Classifier 2 | MeanDiff | Lower | Upper | Adj.p-value |
|---|---|---|---|---|---|---|---|
| GH | CBM | dbCAN_2:HMMER | dbCAN_2:HMMER | 0.093185294 | 0.007568 | 0.178803 | 0.011268 |
| GT | CBM | dbCAN_2:HMMER | dbCAN_2:HMMER | 0.091911891 | 0.006294 | 0.17753 | 0.014707 |
| CBM | AA | dbCAN_3:eCAMI | dbCAN_3:eCAMI | -0.18990795 | -0.28877 | -0.09105 | 2.66E-12 |
| GH | AA | dbCAN_3:eCAMI | dbCAN_3:eCAMI | 0.131449416 | 0.032587 | 0.230312 | 7.16E-05 |
| GT | AA | dbCAN_3:eCAMI | dbCAN_3:eCAMI | 0.13671768 | 0.037855 | 0.235581 | 1.99E-05 |
| CE | CBM | dbCAN_3:eCAMI | dbCAN_3:eCAMI | 0.254043534 | 0.167307 | 0.34078 | 0 |
| GH | CBM | dbCAN_3:eCAMI | dbCAN_3:eCAMI | 0.321357366 | 0.23574 | 0.406975 | 0 |
| GT | CBM | dbCAN_3:eCAMI | dbCAN_3:eCAMI | 0.32662563 | 0.241008 | 0.412243 | 0 |
| PL | CBM | dbCAN_3:eCAMI | dbCAN_3:eCAMI | 0.271826753 | 0.172964 | 0.37069 | 0 |
| GH | CBM | dbCAN_3:HMMER | dbCAN_3:HMMER | 0.094545883 | 0.008928 | 0.180164 | 0.008422 |
| GT | CBM | dbCAN_3:HMMER | dbCAN_3:HMMER | 0.092047602 | 0.00643 | 0.177665 | 0.014299 |
| CBM | AA | CUPP | CUPP | -0.918434718 | -1.0173 | -0.81957 | 0 |
| CE | CBM | CUPP | CUPP | 0.95976874 | 0.873032 | 1.046506 | 0 |
| GH | CBM | CUPP | CUPP | 0.990577189 | 0.904959 | 1.076195 | 0 |
| GT | CBM | CUPP | CUPP | 0.988306849 | 0.902689 | 1.073925 | 0 |
| PL | CBM | CUPP | CUPP | 0.94964539 | 0.850127 | 1.049163 | 0 |
| CBM | AA | dbCAN_2:Hotpep | dbCAN_2:Hotpep | -0.428702977 | -0.52757 | -0.32984 | 0 |
| CE | CBM | dbCAN_2:Hotpep | dbCAN_2:Hotpep | 0.44314086 | 0.356404 | 0.529878 | 0 |
| GH | CBM | dbCAN_2:Hotpep | dbCAN_2:Hotpep | 0.509945492 | 0.424328 | 0.595563 | 0 |
| GT | CBM | dbCAN_2:Hotpep | dbCAN_2:Hotpep | 0.513984388 | 0.428367 | 0.599602 | 0 |
| PL | CBM | dbCAN_2:Hotpep | dbCAN_2:Hotpep | 0.480987281 | 0.381469 | 0.580505 | 0 |

## 5.6 Sensitivity

### SI Figure 22: Sensitivity of binary CAZy class classification performance per CAZy class

Sensitivity was calculated for each test set per CAZy class. The mean sensitivity was then calculated across all test sets and 95% confidence interval (CI) was calculated for each CAZy class, and is plotted in SI figure 22.

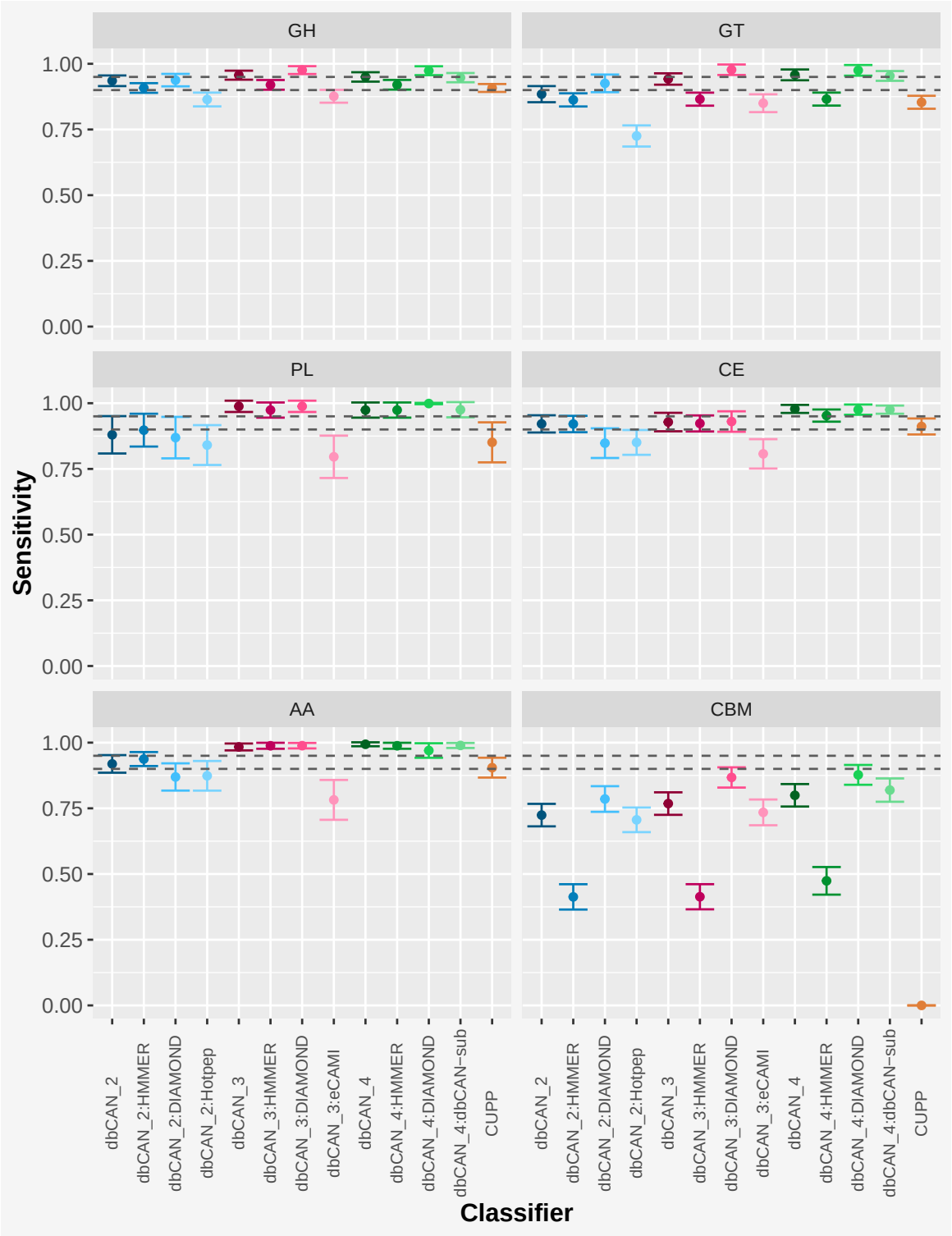



Figure 22: The mean sensitivity and 95% confidence interval (CI) of binary CAZy class classification per CAZy class, across all test sets. Where a tool is incorporated into a CAZyme classifier, the tool name is prefixed with the respective CAZyme classifier. dbCAN_2 and associated tools are coloured blue; dbCAN_3 pink; dbCAN_4 green; and CUPP orange.

## SI Figure 23: Sensitivity of binary CAZy class classification performance per test set

Sensitivity was calculated for each test set per CAZy class, and was plotted in SI figure 23.

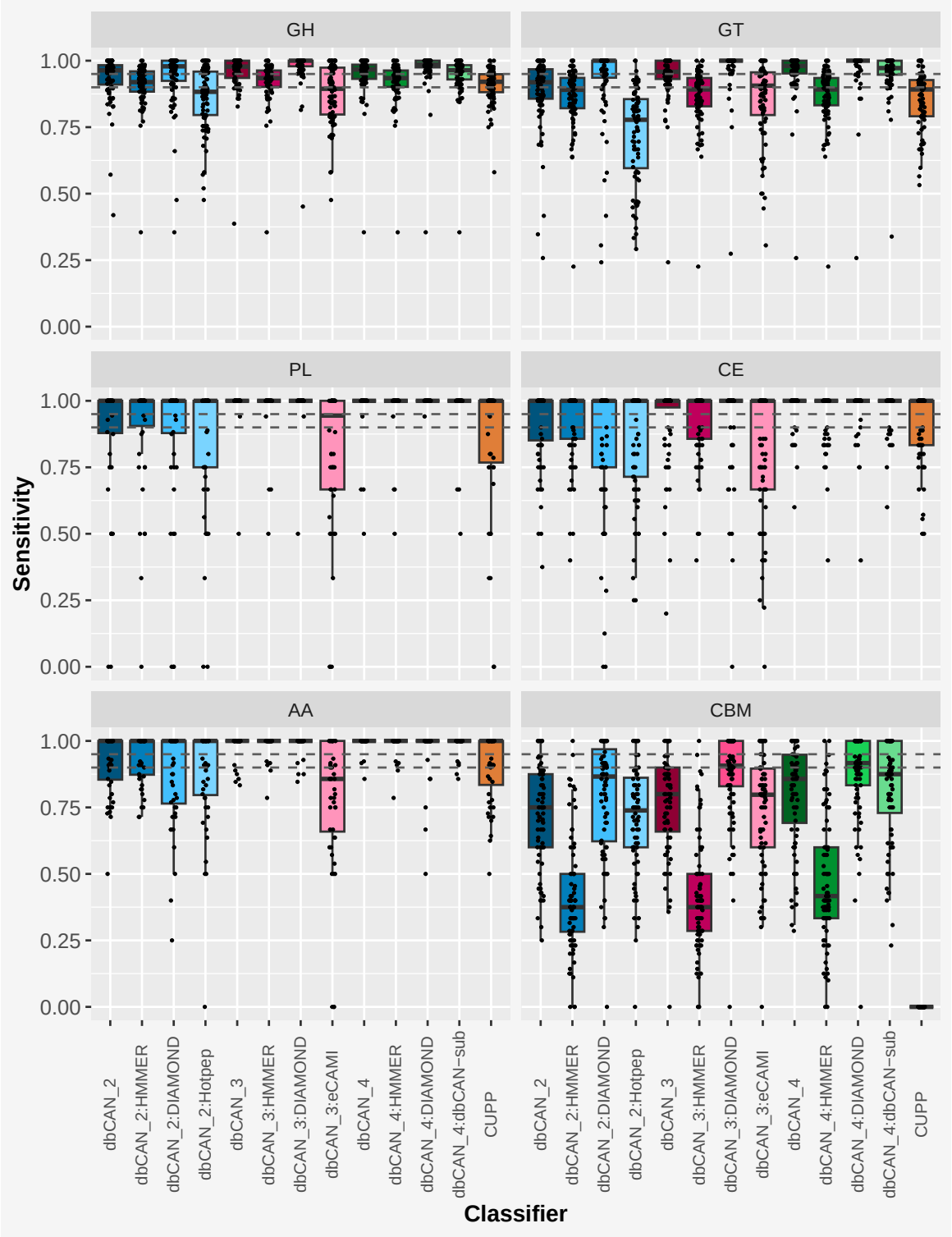



Figure 23: The sensitivity of the binary classification of each CAZy class was calculated per test set, and is plotted as a one-dimensional scatter plot overlaying a box and whisker plot. Where a tool is incorporated into a CAZyme classifier, the tool name is prefixed with the respective CAZyme classifier. dbCAN_2 and associated tools are coloured blue; dbCAN_3 pink; dbCAN_4 green; and CUPP orange.

## SI Table 30: Output of Tukey HSD test for statistically significant differences between the mean sensitivity of CAZy class and classifiers, reporting only tests with a p-value <0.05.

Specifically, only hits where a statistically significant difference (p-value<0.05) is found between the same tool (i.e. Classifier 1 and Classifier 2 are the same) for different CAZy classes (i.e. Class 1 and Class 2 are different). The complete dataset is available in the online repository. Here the pair of CAZy classes, pair of classifiers/tools, the mean difference (Mean Diff), lower and upper 95% confidence interval and the adjusted p-value (Adj.p-value) are reported.

| Class 1 | Class 2 | Classifier 1 | Classifier 2 | MeanDiff | Lower | Upper | Adj.p-value |
|---|---|---|---|---|---|---|---|
| CBM | AA | dbCAN_2:HMMER | dbCAN_2:HMMER | -0.52454 | -0.63788 | -0.41121 | 0 |
| CE | CBM | dbCAN_2:HMMER | dbCAN_2:HMMER | 0.507839 | 0.408406 | 0.607272 | 0 |
| GH | CBM | dbCAN_2:HMMER | dbCAN_2:HMMER | 0.495024 | 0.396874 | 0.593175 | 0 |
| GT | CBM | dbCAN_2:HMMER | dbCAN_2:HMMER | 0.449724 | 0.351574 | 0.547874 | 0 |
| PL | CBM | dbCAN_2:HMMER | dbCAN_2:HMMER | 0.484507 | 0.370422 | 0.598593 | 0 |
| CBM | AA | dbCAN_3:DIAMOND | dbCAN_3:DIAMOND | -0.12087 | -0.23421 | -0.00754 | 0.016636 |
| GH | CBM | dbCAN_3:DIAMOND | dbCAN_3:DIAMOND | 0.108528 | 0.010378 | 0.206678 | 0.008198 |
| GT | CBM | dbCAN_3:DIAMOND | dbCAN_3:DIAMOND | 0.109888 | 0.011738 | 0.208038 | 0.006323 |
| PL | CBM | dbCAN_3:DIAMOND | dbCAN_3:DIAMOND | 0.12062 | 0.006535 | 0.234706 | 0.019532 |
| GH | CBM | dbCAN_3:eCAMI | dbCAN_3:eCAMI | 0.141787 | 0.043636 | 0.239937 | 4.21E-06 |
| GT | CBM | dbCAN_3:eCAMI | dbCAN_3:eCAMI | 0.115428 | 0.017278 | 0.213579 | 0.00209 |
| CBM | AA | dbCAN_3:HMMER | dbCAN_3:HMMER | -0.5745 | -0.68783 | -0.46116 | 0 |
| GT | AA | dbCAN_3:HMMER | dbCAN_3:HMMER | -0.12261 | -0.23594 | -0.00928 | 0.012677 |
| CE | CBM | dbCAN_3:HMMER | dbCAN_3:HMMER | 0.50949 | 0.410057 | 0.608924 | 0 |
| GH | CBM | dbCAN_3:HMMER | dbCAN_3:HMMER | 0.506322 | 0.408171 | 0.604472 | 0 |
| GT | CBM | dbCAN_3:HMMER | dbCAN_3:HMMER | 0.451888 | 0.353738 | 0.550039 | 0 |
| PL | CBM | dbCAN_3:HMMER | dbCAN_3:HMMER | 0.560437 | 0.446351 | 0.674522 | 0 |
| CBM | AA | dbCAN_2 | dbCAN_2 | -0.19481 | -0.30814 | -0.08147 | 1.87E-09 |
| CE | CBM | dbCAN_2 | dbCAN_2 | 0.197116 | 0.097683 | 0.29655 | 7.13E-14 |
| GH | CBM | dbCAN_2 | dbCAN_2 | 0.211198 | 0.113047 | 0.309348 | 0 |
| GT | CBM | dbCAN_2 | dbCAN_2 | 0.160321 | 0.06217 | 0.258471 | 2.38E-08 |
| PL | CBM | dbCAN_2 | dbCAN_2 | 0.155548 | 0.041462 | 0.269633 | 3.19E-05 |
| PL | CBM | dbCAN_4:DIAMOND | dbCAN_4:DIAMOND | 0.121415 | 0.007329 | 0.2355 | 0.01731 |
| GH | CBM | dbCAN_2:DIAMOND | dbCAN_2:DIAMOND | 0.152433 | 0.054283 | 0.250583 | 2.32E-07 |
| GT | CBM | dbCAN_2:DIAMOND | dbCAN_2:DIAMOND | 0.140075 | 0.041925 | 0.238225 | 6.57E-06 |
| CBM | AA | dbCAN_4 | dbCAN_4 | -0.1942 | -0.30753 | -0.08087 | 2.20E-09 |
| CE | CBM | dbCAN_4 | dbCAN_4 | 0.178769 | 0.079335 | 0.278202 | 1.58E-10 |
| GH | CBM | dbCAN_4 | dbCAN_4 | 0.150546 | 0.052396 | 0.248697 | 3.94E-07 |
| GT | CBM | dbCAN_4 | dbCAN_4 | 0.158334 | 0.060183 | 0.256484 | 4.27E-08 |
| PL | CBM | dbCAN_4 | dbCAN_4 | 0.174441 | 0.060356 | 0.288527 | 4.49E-07 |
| CBM | AA | CUPP | CUPP | -0.90473 | -1.01806 | -0.7914 | 0 |
| CE | CBM | CUPP | CUPP | 0.91143 | 0.811996 | 1.010863 | 0 |
| GH | CBM | CUPP | CUPP | 0.908001 | 0.809851 | 1.006151 | 0 |
| GT | CBM | CUPP | CUPP | 0.853565 | 0.755415 | 0.951716 | 0 |
| PL | CBM | CUPP | CUPP | 0.851075 | 0.736989 | 0.96516 | 0 |
| CBM | AA | dbCAN_4:sub | dbCAN_4:sub | -0.16986 | -0.2832 | -0.05653 | 1.02E-06 |
| CE | CBM | dbCAN_4:sub | dbCAN_4:sub | 0.156137 | 0.056703 | 0.25557 | 1.44E-07 |
| GH | CBM | dbCAN_4:sub | dbCAN_4:sub | 0.127935 | 0.029785 | 0.226085 | 0.000132 |
| GT | CBM | dbCAN_4:sub | dbCAN_4:sub | 0.134487 | 0.036336 | 0.232637 | 2.71E-05 |
| PL | CBM | dbCAN_4:sub | dbCAN_4:sub | 0.155855 | 0.041769 | 0.26994 | 2.99E-05 |
| CBM | AA | dbCAN_3 | dbCAN_3 | -0.21553 | -0.32886 | -0.10219 | 5.24E-12 |
| CE | CBM | dbCAN_3 | dbCAN_3 | 0.160267 | 0.060834 | 0.259701 | 4.44E-08 |
| GH | CBM | dbCAN_3 | dbCAN_3 | 0.188714 | 0.090563 | 0.286864 | 2.50E-12 |
| GT | CBM | dbCAN_3 | dbCAN_3 | 0.174102 | 0.075952 | 0.272253 | 3.39E-10 |
| PL | CBM | dbCAN_3 | dbCAN_3 | 0.220104 | 0.106018 | 0.334189 | 1.96E-12 |
| CBM | AA | dbCAN_2:Hotpep | dbCAN_2:Hotpep | -0.16759 | -0.28092 | -0.05425 | 1.73E-06 |
| GT | AA | dbCAN_2:Hotpep | dbCAN_2:Hotpep | -0.14831 | -0.26165 | -0.03498 | 0.000117 |
| CE | CBM | dbCAN_2:Hotpep | dbCAN_2:Hotpep | 0.144679 | 0.045245 | 0.244112 | 3.21E-06 |
| GH | CBM | dbCAN_2:Hotpep | dbCAN_2:Hotpep | 0.157965 | 0.059814 | 0.256115 | 4.75E-08 |
| PL | CBM | dbCAN_2:Hotpep | dbCAN_2:Hotpep | 0.134608 | 0.020522 | 0.248693 | 0.001933 |
| GT | CE | dbCAN_2:Hotpep | dbCAN_2:Hotpep | -0.1254 | -0.22484 | -0.02597 | 0.000343 |
| GT | GH | dbCAN_2:Hotpep | dbCAN_2:Hotpep | -0.13869 | -0.23684 | -0.04054 | 9.39E-06 |
| PL | GT | dbCAN_2:Hotpep | dbCAN_2:Hotpep | 0.115333 | 0.001247 | 0.229418 | 0.042111 |
| CBM | AA | dbCAN_4:HMMER | dbCAN_4:HMMER | -0.51395 | -0.62728 | -0.40061 | 0 |
| GT | AA | dbCAN_4:HMMER | dbCAN_4:HMMER | -0.1223 | -0.23563 | -0.00896 | 0.013317 |
| CE | CBM | dbCAN_4:HMMER | dbCAN_4:HMMER | 0.478853 | 0.379419 | 0.578286 | 0 |
| GH | CBM | dbCAN_4:HMMER | dbCAN_4:HMMER | 0.446133 | 0.347983 | 0.544284 | 0 |
| GT | CBM | dbCAN_4:HMMER | dbCAN_4:HMMER | 0.391649 | 0.293499 | 0.4898 | 0 |
| PL | CBM | dbCAN_4:HMMER | dbCAN_4:HMMER | 0.499886 | 0.3858 | 0.613971 | 0 |

# 6 Performance per CAZy class

## 6.1 Glycoside Hydrolases

**SI table 31: Output of Tukey HSD test for statistically significant differences between the mean specificity of tools classifying GH CAZyme domains (overleaf)**

| Classifier 1 | Classifier 2 | Mean Difference | Lower 95% Confidence Interval | Upper 95% Confidence Interval | Adjusted P-value |
|---|---|---|---|---|---|
| dbCAN_3:HMMER | dbCAN_2:Hotpep | 0.010567864 | -0.002776047 | 0.023911775 | 0.293625656 |
| dbCAN_4:HMMER | dbCAN_2:Hotpep | 0.010567864 | -0.002776047 | 0.023911775 | 0.293625656 |
| CUPP | dbCAN_2:Hotpep | 0.009990176 | -0.003353735 | 0.023334088 | 0.384841429 |
| dbCAN_3:DIAMOND | dbCAN_3:HMMER | -0.009444583 | -0.022788494 | 0.003899328 | 0.48056968 |
| dbCAN_4:HMMER | dbCAN_3:DIAMOND | 0.009444583 | -0.003899328 | 0.022788494 | 0.48056968 |
| dbCAN_2:Hotpep | dbCAN_2:HMMER | -0.009405052 | -0.022748963 | 0.003938859 | 0.487745322 |
| dbCAN_4 | dbCAN_2:Hotpep | 0.009369719 | -0.003974192 | 0.02271363 | 0.494177379 |
| dbCAN_4:dbCANsub | dbCAN_2:Hotpep | 0.009111072 | -0.00423284 | 0.022454983 | 0.541635971 |
| dbCAN_3:eCAMI | dbCAN_3:HMMER | -0.009105877 | -0.022449788 | 0.004238034 | 0.542593301 |
| dbCAN_4:HMMER | dbCAN_3:eCAMI | 0.009105877 | -0.004238034 | 0.022449788 | 0.542593301 |
| dbCAN_3 | dbCAN_2:Hotpep | 0.009049088 | -0.004294823 | 0.022392999 | 0.553062476 |
| CUPP | dbCAN_3:DIAMOND | 0.008866895 | -0.004477016 | 0.022210807 | 0.586622248 |
| CUPP | dbCAN_3:eCAMI | 0.008528189 | -0.004815722 | 0.021872101 | 0.648211387 |
| dbCAN_2:Hotpep | dbCAN_2 | -0.008412168 | -0.021756079 | 0.004931743 | 0.668840846 |
| dbCAN_3:DIAMOND | dbCAN_2:HMMER | -0.008281771 | -0.021625683 | 0.00506214 | 0.691615801 |
| dbCAN_4 | dbCAN_3:DIAMOND | 0.008246438 | -0.005097473 | 0.02159035 | 0.697701528 |
| dbCAN_4:dbCANsub | dbCAN_3:DIAMOND | 0.007987791 | -0.00535612 | 0.021331702 | 0.740940176 |
| dbCAN_3:eCAMI | dbCAN_2:HMMER | -0.007943065 | -0.021286976 | 0.005400846 | 0.74815608 |
| dbCAN_3:DIAMOND | dbCAN_3 | -0.007925807 | -0.021269718 | 0.005418104 | 0.750918012 |
| dbCAN_4 | dbCAN_3:eCAMI | 0.007907732 | -0.005436179 | 0.021251643 | 0.753796877 |
| dbCAN_4:dbCANsub | dbCAN_3:eCAMI | 0.007649085 | -0.005694826 | 0.020992996 | 0.793352509 |
| dbCAN_3:HMMER | dbCAN_2:DIAMOND | 0.00762991 | -0.005714001 | 0.020973821 | 0.79615551 |
| dbCAN_4:HMMER | dbCAN_2:DIAMOND | 0.00762991 | -0.005714001 | 0.020973821 | 0.79615551 |
| dbCAN_3:eCAMI | dbCAN_3 | -0.007587101 | -0.020931012 | 0.00575681 | 0.802345171 |
| dbCAN_4:DIAMOND | dbCAN_3:HMMER | -0.007419494 | -0.020763406 | 0.005924417 | 0.825643477 |
| dbCAN_4:DIAMOND | dbCAN_4:HMMER | -0.007419494 | -0.020763406 | 0.005924417 | 0.825643477 |
| dbCAN_3:DIAMOND | dbCAN_2 | -0.007288887 | -0.020632798 | 0.006055024 | 0.842722167 |
| CUPP | dbCAN_2:DIAMOND | 0.007052222 | -0.006291689 | 0.020396133 | 0.871151333 |
| dbCAN_3:eCAMI | dbCAN_2 | -0.006950181 | -0.020294092 | 0.00639373 | 0.882381606 |
| CUPP | dbCAN_4:DIAMOND | 0.006841807 | -0.006502104 | 0.020185718 | 0.893622793 |
| dbCAN_2:DIAMOND | dbCAN_2:HMMER | -0.006467098 | -0.019811009 | 0.006876813 | 0.927077056 |
| dbCAN_4 | dbCAN_2:DIAMOND | 0.006431765 | -0.006912146 | 0.019775676 | 0.929806808 |
| dbCAN_4:DIAMOND | dbCAN_2:HMMER | -0.006256683 | -0.019600594 | 0.007087228 | 0.942295195 |
| dbCAN_4:DIAMOND | dbCAN_4 | -0.00622135 | -0.019565261 | 0.007122561 | 0.944610765 |
| dbCAN_4:dbCANsub | dbCAN_2:DIAMOND | 0.006173117 | -0.007170794 | 0.019517029 | 0.947663765 |
| dbCAN_3 | dbCAN_2:DIAMOND | 0.006111134 | -0.007232778 | 0.019455045 | 0.95140744 |
| dbCAN_4:dbCANsub | dbCAN_4:DIAMOND | 0.005962702 | -0.007381209 | 0.019306613 | 0.959578253 |
| dbCAN_4:DIAMOND | dbCAN_3 | -0.005900718 | -0.01924463 | 0.007443193 | 0.962671701 |
| dbCAN_2:DIAMOND | dbCAN_2 | -0.005474214 | -0.018818125 | 0.007869698 | 0.979402457 |
| dbCAN_4:DIAMOND | dbCAN_2 | -0.005263798 | -0.01860771 | 0.008080113 | 0.98512903 |
| dbCAN_4:DIAMOND | dbCAN_2:Hotpep | 0.00314837 | -0.010195542 | 0.016492281 | 0.999897642 |
| dbCAN_2:Hotpep | dbCAN_2:DIAMOND | -0.002937954 | -0.016281865 | 0.010405957 | 0.999951212 |
| dbCAN_3:HMMER | dbCAN_2 | 0.002155696 | -0.011188215 | 0.015499607 | 0.999998428 |
| dbCAN_4:HMMER | dbCAN_2 | 0.002155696 | -0.011188215 | 0.015499607 | 0.999998428 |
| dbCAN_4:DIAMOND | dbCAN_3:DIAMOND | 0.002025089 | -0.011318823 | 0.015369 | 0.999999228 |
| dbCAN_3:DIAMOND | dbCAN_2:DIAMOND | -0.001814673 | -0.015158584 | 0.011529238 | 0.999999781 |
| dbCAN_4:DIAMOND | dbCAN_3:eCAMI | 0.001686383 | -0.011657529 | 0.015030294 | 0.999999906 |
| CUPP | dbCAN_2 | 0.001578008 | -0.011765903 | 0.01492192 | 0.999999957 |
| dbCAN_3:HMMER | dbCAN_3 | 0.001518776 | -0.011825135 | 0.014862687 | 0.999999972 |
| dbCAN_4:HMMER | dbCAN_3 | 0.001518776 | -0.011825135 | 0.014862687 | 0.999999972 |
| dbCAN_3:eCAMI | dbCAN_2:DIAMOND | -0.001475967 | -0.014819878 | 0.011867944 | 0.99999998 |
| dbCAN_3:eCAMI | dbCAN_2:Hotpep | 0.001461987 | -0.011881924 | 0.014805898 | 0.999999982 |
| dbCAN_4:dbCANsub | dbCAN_3:HMMER | -0.001456792 | -0.014800703 | 0.011887119 | 0.999999983 |
| dbCAN_4:dbCANsub | dbCAN_4:HMMER | -0.001456792 | -0.014800703 | 0.011887119 | 0.999999983 |
| dbCAN_4 | dbCAN_3:HMMER | -0.001198145 | -0.014542056 | 0.012145766 | 0.999999998 |
| dbCAN_4:HMMER | dbCAN_4 | 0.001198145 | -0.012145766 | 0.014542056 | 0.999999998 |
| dbCAN_3:HMMER | dbCAN_2:HMMER | 0.001162812 | -0.0121811 | 0.014506723 | 0.999999999 |
| dbCAN_4:HMMER | dbCAN_2:HMMER | 0.001162812 | -0.0121811 | 0.014506723 | 0.999999999 |
| dbCAN_3:DIAMOND | dbCAN_2:Hotpep | 0.001123281 | -0.01222063 | 0.014467192 | 0.999999999 |
| dbCAN_2:HMMER | dbCAN_2 | 0.000992884 | -0.012351027 | 0.014336796 | 1 |
| dbCAN_4 | dbCAN_2 | 0.000957551 | -0.01238636 | 0.014301463 | 1 |
| CUPP | dbCAN_3 | 0.000941088 | -0.012402823 | 0.014285 | 1 |
| CUPP | dbCAN_4:dbCANsub | 0.000879105 | -0.012464806 | 0.014223016 | 1 |
| dbCAN_4:dbCANsub | dbCAN_2 | 0.000698904 | -0.012645007 | 0.014042815 | 1 |
| dbCAN_3 | dbCAN_2 | 0.00063692 | -0.012706991 | 0.013980831 | 1 |
| CUPP | dbCAN_4 | 0.000620457 | -0.012723454 | 0.013964368 | 1 |
| dbCAN_3 | dbCAN_2:HMMER | -0.000355964 | -0.013699876 | 0.012987947 | 1 |
| dbCAN_4 | dbCAN_2:HMMER | -3.53E-05 | -0.013379244 | 0.013308578 | 1 |
| dbCAN_4:dbCANsub | dbCAN_2:HMMER | -0.000293981 | -0.013637892 | 0.013049931 | 1 |
| CUPP | dbCAN_2:HMMER | 0.000585124 | -0.012758787 | 0.013929035 | 1 |
| dbCAN_4:DIAMOND | dbCAN_2:DIAMOND | 0.000210415 | -0.013133496 | 0.013554326 | 1 |
| dbCAN_4 | dbCAN_3 | 0.000320631 | -0.01302328 | 0.013664543 | 1 |
| dbCAN_4:dbCANsub | dbCAN_3 | 6.20E-05 | -0.013281927 | 0.013405895 | 1 |
| dbCAN_4:HMMER | dbCAN_3:HMMER | -1.11E-16 | -0.013343911 | 0.013343911 | 1 |
| CUPP | dbCAN_3:HMMER | -0.000577688 | -0.013921599 | 0.012766224 | 1 |
| dbCAN_3:eCAMI | dbCAN_3:DIAMOND | 0.000338706 | -0.013005205 | 0.013682617 | 1 |
| dbCAN_4:dbCANsub | dbCAN_4 | -0.000258648 | -0.013602559 | 0.013085264 | 1 |
| CUPP | dbCAN_4:HMMER | -0.000577688 | -0.013921599 | 0.012766224 | 1 |

**SI table 32: Output of Tukey HSD test for statistically significant differences between the mean F1-score of tools classifying GH CAZyme domains (overleaf)**

| Classifier 1 | Classifier 2 | Mean Difference | Lower 95% Confidence Interval | Upper 95% Confidence Interval | Adjusted P-value |
|---|---|---|---|---|---|
| dbCAN_3:DIAMOND | dbCAN_2:Hotpep | 0.063951631 | 0.031958703 | 0.095944558 | 4.01E-09 |
| dbCAN_4:DIAMOND | dbCAN_2:Hotpep | 0.062763479 | 0.030770551 | 0.094756406 | 8.88E-09 |
| dbCAN_3 | dbCAN_2:Hotpep | 0.057131405 | 0.025138478 | 0.089124333 | 3.21E-07 |
| dbCAN_3:eCAMI | dbCAN_3:DIAMOND | -0.055228367 | -0.087221294 | -0.023235439 | 1.01E-06 |
| dbCAN_4:DIAMOND | dbCAN_3:eCAMI | 0.054040215 | 0.022047288 | 0.086033143 | 2.02E-06 |
| dbCAN_4 | dbCAN_2:Hotpep | 0.053428859 | 0.021435931 | 0.085421786 | 2.87E-06 |
| dbCAN_4:dbCANsub | dbCAN_2:Hotpep | 0.051927297 | 0.019934369 | 0.083920224 | 6.72E-06 |
| dbCAN_3:eCAMI | dbCAN_3 | -0.048408141 | -0.080401069 | -0.016415214 | 4.50E-05 |
| dbCAN_4 | dbCAN_3:eCAMI | 0.044705595 | 0.012712667 | 0.076698522 | 0.000289174 |
| dbCAN_2:Hotpep | dbCAN_2 | -0.044658263 | -0.07665119 | -0.012665335 | 0.000295848 |
| dbCAN_4:dbCANsub | dbCAN_3:eCAMI | 0.043204033 | 0.011211105 | 0.07519696 | 0.000589155 |
| dbCAN_2:Hotpep | dbCAN_2:DIAMOND | -0.042654162 | -0.07464709 | -0.010661235 | 0.000759691 |
| dbCAN_4:HMMER | dbCAN_2:Hotpep | 0.038008681 | 0.006015753 | 0.070001608 | 0.005641619 |
| dbCAN_3:HMMER | dbCAN_2:Hotpep | 0.037804526 | 0.005811599 | 0.069797454 | 0.006123887 |
| dbCAN_3:eCAMI | dbCAN_2 | -0.035934999 | -0.067927927 | -0.003942071 | 0.012656875 |
| dbCAN_3:eCAMI | dbCAN_2:DIAMOND | -0.033930898 | -0.065923826 | -0.001937971 | 0.026160372 |
| dbCAN_3:DIAMOND | dbCAN_2:HMMER | 0.033200621 | 0.001207693 | 0.065193548 | 0.033608753 |
| CUPP | dbCAN_2:Hotpep | 0.03243999 | 0.000447062 | 0.064432917 | 0.043271588 |
| dbCAN_4:DIAMOND | dbCAN_2:HMMER | 0.032012469 | 1.95E-05 | 0.064005396 | 0.049688245 |
| CUPP | dbCAN_3:DIAMOND | -0.031511641 | -0.063504568 | 0.000481287 | 0.058220741 |
| dbCAN_2:Hotpep | dbCAN_2:HMMER | -0.03075101 | -0.062743938 | 0.001241917 | 0.073519569 |
| CUPP | dbCAN_4:DIAMOND | -0.030323489 | -0.062316417 | 0.001669438 | 0.083488855 |
| dbCAN_4:HMMER | dbCAN_3:eCAMI | 0.029285417 | -0.002707511 | 0.061278344 | 0.112313034 |
| dbCAN_3:eCAMI | dbCAN_3:HMMER | -0.029081263 | -0.06107419 | 0.002911665 | 0.118812725 |
| dbCAN_3 | dbCAN_2:HMMER | 0.026380395 | -0.005612533 | 0.058373323 | 0.23395358 |
| dbCAN_3:DIAMOND | dbCAN_3:HMMER | 0.026147104 | -0.005845823 | 0.058140032 | 0.246569281 |
| dbCAN_4:HMMER | dbCAN_3:DIAMOND | -0.02594295 | -0.057935878 | 0.006049977 | 0.257957966 |
| dbCAN_4:DIAMOND | dbCAN_3:HMMER | 0.024958953 | -0.007033975 | 0.05695188 | 0.317264631 |
| dbCAN_4:DIAMOND | dbCAN_4:HMMER | 0.024754798 | -0.007238129 | 0.056747726 | 0.330440681 |
| CUPP | dbCAN_3 | -0.024691415 | -0.056684343 | 0.007301512 | 0.334588715 |
| CUPP | dbCAN_3:eCAMI | 0.023716726 | -0.008276202 | 0.055709653 | 0.401503637 |
| dbCAN_4 | dbCAN_2:HMMER | 0.022677849 | -0.009315079 | 0.054670776 | 0.478014548 |
| dbCAN_3:eCAMI | dbCAN_2:HMMER | -0.022027746 | -0.054020674 | 0.009965181 | 0.527555383 |
| dbCAN_3:DIAMOND | dbCAN_2:DIAMOND | 0.021297468 | -0.010695459 | 0.053290396 | 0.583670962 |
| dbCAN_4:dbCANsub | dbCAN_2:HMMER | 0.021176287 | -0.010816641 | 0.053169214 | 0.592956941 |
| CUPP | dbCAN_4 | -0.020988869 | -0.052981796 | 0.011004059 | 0.607274619 |
| dbCAN_4:DIAMOND | dbCAN_2:DIAMOND | 0.020109317 | -0.011883611 | 0.052102244 | 0.673205605 |
| CUPP | dbCAN_4:dbCANsub | -0.019487307 | -0.051480234 | 0.012505621 | 0.717811216 |
| dbCAN_3:HMMER | dbCAN_3 | -0.019326879 | -0.051319806 | 0.012666049 | 0.728947537 |
| dbCAN_3:DIAMOND | dbCAN_2 | 0.019293368 | -0.01269956 | 0.051286295 | 0.731252352 |
| dbCAN_4:HMMER | dbCAN_3 | -0.019122725 | -0.051115652 | 0.012870203 | 0.742868998 |
| dbCAN_4:DIAMOND | dbCAN_2 | 0.018105216 | -0.013887711 | 0.050098144 | 0.807420788 |
| dbCAN_4 | dbCAN_3:HMMER | 0.015624332 | -0.016368595 | 0.04761726 | 0.923120131 |
| dbCAN_4:HMMER | dbCAN_4 | -0.015420178 | -0.047413106 | 0.016572749 | 0.929819922 |
| dbCAN_3 | dbCAN_2:DIAMOND | 0.014477243 | -0.017515685 | 0.04647017 | 0.955553856 |
| dbCAN_4:dbCANsub | dbCAN_3:HMMER | 0.01412277 | -0.017870157 | 0.046115698 | 0.963166293 |
| dbCAN_4:dbCANsub | dbCAN_4:HMMER | 0.013918616 | -0.018074311 | 0.045911544 | 0.967084383 |
| dbCAN_2:HMMER | dbCAN_2 | -0.013907253 | -0.04590018 | 0.018085675 | 0.967292863 |
| dbCAN_3 | dbCAN_2 | 0.012473142 | -0.019519785 | 0.04446607 | 0.98653596 |
| CUPP | dbCAN_2 | -0.012218273 | -0.044211201 | 0.019774654 | 0.988719882 |
| dbCAN_4:dbCANsub | dbCAN_3:DIAMOND | -0.012024334 | -0.044017261 | 0.019968594 | 0.990182933 |
| dbCAN_2:DIAMOND | dbCAN_2:HMMER | 0.011903152 | -0.020089775 | 0.04389608 | 0.991016381 |
| dbCAN_4:dbCANsub | dbCAN_4:DIAMOND | -0.010836182 | -0.04282911 | 0.021156745 | 0.996158136 |
| dbCAN_4 | dbCAN_2:DIAMOND | 0.010774697 | -0.021218231 | 0.042767624 | 0.996356235 |
| dbCAN_4 | dbCAN_3:DIAMOND | -0.010522772 | -0.042515699 | 0.021470156 | 0.997081066 |
| CUPP | dbCAN_2:DIAMOND | -0.010214172 | -0.0422071 | 0.021778755 | 0.997800054 |
| dbCAN_4:DIAMOND | dbCAN_4 | 0.00933462 | -0.022658307 | 0.041327548 | 0.999086498 |
| dbCAN_4:dbCANsub | dbCAN_2:DIAMOND | 0.009273135 | -0.022719793 | 0.041266062 | 0.999144667 |
| dbCAN_4 | dbCAN_2 | 0.008770596 | -0.023222332 | 0.040763523 | 0.999511979 |
| dbCAN_3:eCAMI | dbCAN_2:Hotpep | 0.008723264 | -0.023269664 | 0.040716191 | 0.999538136 |
| dbCAN_4:dbCANsub | dbCAN_2 | 0.007269034 | -0.024723894 | 0.039261961 | 0.999931563 |
| dbCAN_4:HMMER | dbCAN_2:HMMER | 0.00725767 | -0.024735257 | 0.039250598 | 0.999932701 |
| dbCAN_3:HMMER | dbCAN_2:HMMER | 0.007053516 | -0.024939411 | 0.039046444 | 0.999950491 |
| dbCAN_3:HMMER | dbCAN_2 | -0.006853736 | -0.038846664 | 0.025139191 | 0.999963726 |
| dbCAN_3:DIAMOND | dbCAN_3 | 0.006820226 | -0.025172702 | 0.038813153 | 0.999965607 |
| dbCAN_4:HMMER | dbCAN_2 | -0.006649582 | -0.03864251 | 0.025343345 | 0.999973904 |
| dbCAN_4:DIAMOND | dbCAN_3 | 0.005632074 | -0.026360854 | 0.037625001 | 0.999995856 |
| CUPP | dbCAN_4:HMMER | -0.005568691 | -0.037561618 | 0.026424237 | 0.99999635 |
| CUPP | dbCAN_3:HMMER | -0.005364537 | -0.037357464 | 0.026628391 | 0.999997604 |
| dbCAN_4:dbCANsub | dbCAN_3 | -0.005204108 | -0.037197036 | 0.026788819 | 0.9999983 |
| dbCAN_3:HMMER | dbCAN_2:DIAMOND | -0.004849636 | -0.036842563 | 0.027143292 | 0.999999238 |
| dbCAN_4:HMMER | dbCAN_2:DIAMOND | -0.004645482 | -0.036638409 | 0.027347446 | 0.999999534 |
| dbCAN_4 | dbCAN_3 | -0.003702546 | -0.035695474 | 0.028290381 | 0.999999966 |
| dbCAN_2:DIAMOND | dbCAN_2 | -0.002004101 | -0.033997028 | 0.029988827 | 1 |
| CUPP | dbCAN_2:HMMER | 0.00168898 | -0.030303948 | 0.033681907 | 1 |
| dbCAN_4:dbCANsub | dbCAN_4 | -0.001501562 | -0.03349449 | 0.030491365 | 1 |
| dbCAN_4:HMMER | dbCAN_3:HMMER | 0.000204154 | -0.031788773 | 0.032197082 | 1 |
| dbCAN_4:DIAMOND | dbCAN_3:DIAMOND | -0.001188152 | -0.033181079 | 0.030804776 | 1 |

**SI table 33: Output of Tukey HSD test for statistically significant differences between the mean accuracy of tools classifying GH CAZyme domains (overleaf)**

| Classifier 1 | Classifier 2 | Mean Difference | Lower 95% Confidence Interval | Upper 95% Confidence Interval | Adjusted P-value |
|---|---|---|---|---|---|
| dbCAN_4:DIAMOND | dbCAN_2:Hotpep | 0.051543228 | 0.032849403 | 0.070237053 | 4.13E-13 |
| dbCAN_3:DIAMOND | dbCAN_2:Hotpep | 0.051527487 | 0.032833662 | 0.070221312 | 4.13E-13 |
| dbCAN_3 | dbCAN_2:Hotpep | 0.047135627 | 0.028441802 | 0.065829452 | 4.35E-13 |
| dbCAN_4:DIAMOND | dbCAN_3:eCAMI | 0.044858011 | 0.026164185 | 0.063551836 | 7.71E-13 |
| dbCAN_3:eCAMI | dbCAN_3:DIAMOND | -0.044842269 | -0.063536094 | -0.026148444 | 7.78E-13 |
| dbCAN_4 | dbCAN_2:Hotpep | 0.044648172 | 0.025954347 | 0.063341997 | 8.69E-13 |
| dbCAN_4:dbCANsub | dbCAN_2:Hotpep | 0.043438702 | 0.024744877 | 0.062132528 | 2.60E-12 |
| dbCAN_3:eCAMI | dbCAN_3 | -0.040450409 | -0.059144235 | -0.021756584 | 1.01E-10 |
| dbCAN_4 | dbCAN_3:eCAMI | 0.037962954 | 0.019269129 | 0.056656779 | 2.01E-09 |
| dbCAN_4:dbCANsub | dbCAN_3:eCAMI | 0.036753485 | 0.01805966 | 0.05544731 | 8.11E-09 |
| dbCAN_2:Hotpep | dbCAN_2 | -0.036086087 | -0.054779912 | -0.017392262 | 1.72E-08 |
| dbCAN_2:Hotpep | dbCAN_2:DIAMOND | -0.034656622 | -0.053350447 | -0.015962797 | 8.29E-08 |
| dbCAN_4:HMMER | dbCAN_2:Hotpep | 0.030428579 | 0.011734754 | 0.049122404 | 6.18E-06 |
| dbCAN_3:HMMER | dbCAN_2:Hotpep | 0.030178579 | 0.011484754 | 0.048872404 | 7.85E-06 |
| dbCAN_3:eCAMI | dbCAN_2 | -0.029400869 | -0.048094694 | -0.010707044 | 1.63E-05 |
| dbCAN_3:eCAMI | dbCAN_2:DIAMOND | -0.027971404 | -0.046665229 | -0.009277579 | 5.95E-05 |
| dbCAN_4:DIAMOND | dbCAN_2:HMMER | 0.027646064 | 0.008952239 | 0.046339889 | 7.92E-05 |
| dbCAN_3:DIAMOND | dbCAN_2:HMMER | 0.027630322 | 0.008936497 | 0.046324148 | 8.03E-05 |
| CUPP | dbCAN_4:DIAMOND | -0.027246722 | -0.045940548 | -0.008552897 | 0.000111969 |
| CUPP | dbCAN_3:DIAMOND | -0.027230981 | -0.045924806 | -0.008537156 | 0.000113497 |
| CUPP | dbCAN_2:Hotpep | 0.024296506 | 0.00560268 | 0.042990331 | 0.001234991 |
| dbCAN_2:Hotpep | dbCAN_2:HMMER | -0.023897164 | -0.04259099 | -0.005203339 | 0.001671313 |
| dbCAN_4:HMMER | dbCAN_3:eCAMI | 0.023743362 | 0.005049537 | 0.042437187 | 0.001875083 |
| dbCAN_3:eCAMI | dbCAN_3:HMMER | -0.023493362 | -0.042187187 | -0.004799537 | 0.002256642 |
| dbCAN_3 | dbCAN_2:HMMER | 0.023238463 | 0.004544637 | 0.041932288 | 0.002719515 |
| CUPP | dbCAN_3 | -0.022839121 | -0.041532946 | -0.004145296 | 0.003625804 |
| dbCAN_4:DIAMOND | dbCAN_3:HMMER | 0.021364649 | 0.002670824 | 0.040058474 | 0.009965512 |
| dbCAN_3:DIAMOND | dbCAN_3:HMMER | 0.021348907 | 0.002655082 | 0.040042733 | 0.01006919 |
| dbCAN_4:DIAMOND | dbCAN_4:HMMER | 0.021114649 | 0.002420824 | 0.039808474 | 0.011732565 |
| dbCAN_4:HMMER | dbCAN_3:DIAMOND | -0.021098907 | -0.039792733 | -0.002405082 | 0.011852812 |
| dbCAN_4 | dbCAN_2:HMMER | 0.020751008 | 0.002057182 | 0.039444833 | 0.014811983 |
| CUPP | dbCAN_4 | -0.020351666 | -0.039045491 | -0.001657841 | 0.019016753 |
| dbCAN_4:dbCANsub | dbCAN_2:HMMER | 0.019541538 | 0.000847713 | 0.038235363 | 0.030945823 |
| CUPP | dbCAN_4:dbCANsub | -0.019142197 | -0.037836022 | -0.000448372 | 0.038943902 |
| CUPP | dbCAN_3:eCAMI | 0.017611288 | -0.001082537 | 0.036305113 | 0.088087082 |
| dbCAN_3:eCAMI | dbCAN_2:HMMER | -0.017211947 | -0.035905772 | 0.001481878 | 0.107074812 |
| dbCAN_3:HMMER | dbCAN_3 | -0.016957048 | -0.035650873 | 0.001736778 | 0.120800285 |
| dbCAN_4:DIAMOND | dbCAN_2:DIAMOND | 0.016886606 | -0.001807219 | 0.035580431 | 0.124825203 |
| dbCAN_3:DIAMOND | dbCAN_2:DIAMOND | 0.016870865 | -0.00182296 | 0.03556469 | 0.125738657 |
| dbCAN_4:HMMER | dbCAN_3 | -0.016707048 | -0.035400873 | 0.001986778 | 0.135553174 |
| dbCAN_4:DIAMOND | dbCAN_2 | 0.015457141 | -0.003236684 | 0.034150966 | 0.230082424 |
| dbCAN_3:DIAMOND | dbCAN_2 | 0.0154414 | -0.003252425 | 0.034135225 | 0.231501887 |
| dbCAN_4 | dbCAN_3:HMMER | 0.014469593 | -0.004224233 | 0.033163418 | 0.329873008 |
| dbCAN_4:HMMER | dbCAN_4 | -0.014219593 | -0.032913418 | 0.004474233 | 0.358355236 |
| dbCAN_4:dbCANsub | dbCAN_3:HMMER | 0.013260123 | -0.005433702 | 0.031953948 | 0.476825846 |
| dbCAN_4:dbCANsub | dbCAN_4:HMMER | 0.013010123 | -0.005683702 | 0.031703948 | 0.509329874 |
| dbCAN_3 | dbCAN_2:DIAMOND | 0.012479005 | -0.00621482 | 0.03117283 | 0.579119866 |
| dbCAN_2:HMMER | dbCAN_2 | -0.012188923 | -0.030882748 | 0.006504903 | 0.617054906 |
| CUPP | dbCAN_2 | -0.011789581 | -0.030483406 | 0.006904244 | 0.668241142 |
| dbCAN_3 | dbCAN_2 | 0.01104954 | -0.007644285 | 0.029743365 | 0.757031831 |
| dbCAN_2:DIAMOND | dbCAN_2:HMMER | 0.010759458 | -0.007934368 | 0.029453283 | 0.788757098 |
| CUPP | dbCAN_2:DIAMOND | -0.010360116 | -0.029053941 | 0.008333709 | 0.8288939 |
| dbCAN_4 | dbCAN_2:DIAMOND | 0.00999155 | -0.008702275 | 0.028685375 | 0.861924787 |
| dbCAN_4:dbCANsub | dbCAN_2:DIAMOND | 0.00878208 | -0.009911745 | 0.027475906 | 0.941487494 |
| dbCAN_4 | dbCAN_2 | 0.008562085 | -0.01013174 | 0.02725591 | 0.951372089 |
| dbCAN_4:dbCANsub | dbCAN_4:DIAMOND | -0.008104526 | -0.026798351 | 0.010589299 | 0.96796536 |
| dbCAN_4:dbCANsub | dbCAN_3:DIAMOND | -0.008088784 | -0.02678261 | 0.010605041 | 0.968448012 |
| dbCAN_4:dbCANsub | dbCAN_2 | 0.007352615 | -0.01134121 | 0.026046441 | 0.985492494 |
| dbCAN_4:DIAMOND | dbCAN_4 | 0.006895056 | -0.011798769 | 0.025588881 | 0.991677742 |
| dbCAN_4 | dbCAN_3:DIAMOND | -0.006879315 | -0.02557314 | 0.01181451 | 0.991844279 |
| dbCAN_3:eCAMI | dbCAN_2:Hotpep | 0.006685218 | -0.012008608 | 0.025379043 | 0.993683684 |
| dbCAN_4:HMMER | dbCAN_2:HMMER | 0.006531415 | -0.01216241 | 0.02522524 | 0.994885832 |
| dbCAN_3:HMMER | dbCAN_2:HMMER | 0.006281415 | -0.01241241 | 0.02497524 | 0.996432993 |
| CUPP | dbCAN_4:HMMER | -0.006132074 | -0.024825899 | 0.012561752 | 0.997154159 |
| dbCAN_3:HMMER | dbCAN_2 | -0.005907508 | -0.024601333 | 0.012786318 | 0.998005668 |
| CUPP | dbCAN_3:HMMER | -0.005882074 | -0.024575899 | 0.012811752 | 0.99808678 |
| dbCAN_4:HMMER | dbCAN_2 | -0.005657508 | -0.024351333 | 0.013036318 | 0.998689286 |
| dbCAN_3:HMMER | dbCAN_2:DIAMOND | -0.004478043 | -0.023171868 | 0.014215783 | 0.999879778 |
| dbCAN_4:DIAMOND | dbCAN_3 | 0.004407601 | -0.014286224 | 0.023101426 | 0.999898386 |
| dbCAN_3:DIAMOND | dbCAN_3 | 0.00439186 | -0.014301965 | 0.023085685 | 0.999902179 |
| dbCAN_4:HMMER | dbCAN_2:DIAMOND | -0.004228043 | -0.022921868 | 0.014465783 | 0.999934832 |
| dbCAN_4:dbCANsub | dbCAN_3 | -0.003696925 | -0.02239075 | 0.014996901 | 0.999984879 |
| dbCAN_4 | dbCAN_3 | -0.002487455 | -0.02118128 | 0.01620637 | 0.99999983 |
| dbCAN_2:DIAMOND | dbCAN_2 | -0.001429465 | -0.02012329 | 0.01726436 | 1 |
| dbCAN_4:dbCANsub | dbCAN_4 | -0.00120947 | -0.019903295 | 0.017484356 | 1 |
| CUPP | dbCAN_2:HMMER | 0.000399341 | -0.018294484 | 0.019093166 | 1 |
| dbCAN_4:HMMER | dbCAN_3:HMMER | 0.00025 | -0.018443825 | 0.018943825 | 1 |
| dbCAN_4:DIAMOND | dbCAN_3:DIAMOND | 1.57E-05 | -0.018678084 | 0.018709566 | 1 |

**SI table 34: Output of Tukey HSD test for statistically significant differences between the mean sensitivity of tools classifying GH CAZyme domains (overleaf)**

| Classifier 1 | Classifier 2 | Mean Difference | Lower 95% Confidence Interval | Upper 95% Confidence Interval | Adjusted P-value |
|---|---|---|---|---|---|
| dbCAN_3:DIAMOND | dbCAN_2:Hotpep | 0.111922484 | 0.065975685 | 0.157869283 | 5.70E-13 |
| dbCAN_4:DIAMOND | dbCAN_2:Hotpep | 0.109572295 | 0.063625496 | 0.155519095 | 9.12E-13 |
| dbCAN_3:eCAMI | dbCAN_3:DIAMOND | -0.09964685 | -0.145593649 | -0.053700051 | 9.00E-11 |
| dbCAN_4:DIAMOND | dbCAN_3:eCAMI | 0.097296661 | 0.051349862 | 0.14324346 | 2.90E-10 |
| dbCAN_3 | dbCAN_2:Hotpep | 0.092624878 | 0.046678078 | 0.138571677 | 2.78E-09 |
| dbCAN_4 | dbCAN_2:Hotpep | 0.085935304 | 0.039988504 | 0.131882103 | 5.95E-08 |
| dbCAN_4:dbCANsub | dbCAN_2:Hotpep | 0.083162531 | 0.037215731 | 0.12910933 | 1.99E-07 |
| dbCAN_3:eCAMI | dbCAN_3 | -0.080349243 | -0.126296043 | -0.034402444 | 6.56E-07 |
| dbCAN_2:Hotpep | dbCAN_2:DIAMOND | -0.073771585 | -0.119718384 | -0.027824785 | 9.17E-06 |
| dbCAN_4 | dbCAN_3:eCAMI | 0.073659669 | 0.02771287 | 0.119606469 | 9.57E-06 |
| dbCAN_2:Hotpep | dbCAN_2 | -0.071278891 | -0.117225691 | -0.025332092 | 2.36E-05 |
| dbCAN_4:dbCANsub | dbCAN_3:eCAMI | 0.070886896 | 0.024940097 | 0.116833696 | 2.73E-05 |
| dbCAN_3:DIAMOND | dbCAN_2:HMMER | 0.068046041 | 0.022099242 | 0.11399284 | 7.65E-05 |
| CUPP | dbCAN_3:DIAMOND | -0.068016644 | -0.113963444 | -0.022069845 | 7.73E-05 |
| dbCAN_4:DIAMOND | dbCAN_2:HMMER | 0.065695852 | 0.019749053 | 0.111642652 | 0.000174178 |
| CUPP | dbCAN_4:DIAMOND | -0.065666456 | -0.111613255 | -0.019719657 | 0.000175947 |
| dbCAN_3:eCAMI | dbCAN_2:DIAMOND | -0.06149595 | -0.10744275 | -0.015549151 | 0.000703923 |
| dbCAN_3:eCAMI | dbCAN_2 | -0.059003257 | -0.104950057 | -0.013056458 | 0.001539935 |
| dbCAN_3:DIAMOND | dbCAN_3:HMMER | 0.056206789 | 0.01025999 | 0.102153588 | 0.003551386 |
| dbCAN_4:HMMER | dbCAN_2:Hotpep | 0.056078014 | 0.010131215 | 0.102024813 | 0.003686587 |
| dbCAN_4:HMMER | dbCAN_3:DIAMOND | -0.05584447 | -0.101791269 | -0.009897671 | 0.003944049 |
| dbCAN_3:HMMER | dbCAN_2:Hotpep | 0.055715695 | 0.009768896 | 0.101662494 | 0.004093047 |
| dbCAN_4:DIAMOND | dbCAN_3:HMMER | 0.0538566 | 0.007909801 | 0.0998034 | 0.006912566 |
| dbCAN_4:DIAMOND | dbCAN_4:HMMER | 0.053494281 | 0.007547482 | 0.099441081 | 0.007636815 |
| dbCAN_3 | dbCAN_2:HMMER | 0.048748435 | 0.002801635 | 0.094695234 | 0.026043804 |
| CUPP | dbCAN_3 | -0.048719038 | -0.094665837 | -0.002772239 | 0.026230165 |
| CUPP | dbCAN_2:Hotpep | 0.04390584 | -0.00204096 | 0.089852639 | 0.077572065 |
| dbCAN_2:Hotpep | dbCAN_2:HMMER | -0.043876443 | -0.089823242 | 0.002070356 | 0.078046404 |
| dbCAN_4:HMMER | dbCAN_3:eCAMI | 0.04380238 | -0.00214442 | 0.089749179 | 0.079252056 |
| dbCAN_3:eCAMI | dbCAN_3:HMMER | -0.043440061 | -0.08938686 | 0.002506738 | 0.085372065 |
| dbCAN_4 | dbCAN_2:HMMER | 0.04205886 | -0.003887939 | 0.08800566 | 0.112302927 |
| CUPP | dbCAN_4 | -0.042029464 | -0.087976263 | 0.003917335 | 0.112941688 |
| dbCAN_3:DIAMOND | dbCAN_2 | 0.040643593 | -0.005303207 | 0.086590392 | 0.146401268 |
| dbCAN_4:dbCANsub | dbCAN_2:HMMER | 0.039286087 | -0.006660712 | 0.085232887 | 0.185897679 |
| CUPP | dbCAN_4:dbCANsub | -0.039256691 | -0.08520349 | 0.006690108 | 0.186829689 |
| dbCAN_4:DIAMOND | dbCAN_2 | 0.038293404 | -0.007653395 | 0.084240203 | 0.219202534 |
| dbCAN_3:DIAMOND | dbCAN_2:DIAMOND | 0.038150899 | -0.0077959 | 0.084097699 | 0.224294225 |
| dbCAN_3:HMMER | dbCAN_3 | -0.036909183 | -0.082855982 | 0.009037617 | 0.271929559 |
| dbCAN_4:HMMER | dbCAN_3 | -0.036546864 | -0.082493663 | 0.009399936 | 0.28691143 |
| dbCAN_4:DIAMOND | dbCAN_2:DIAMOND | 0.035800711 | -0.010146088 | 0.08174751 | 0.319233145 |
| CUPP | dbCAN_3:eCAMI | 0.031630205 | -0.014316594 | 0.077577005 | 0.527825174 |
| dbCAN_3:eCAMI | dbCAN_2:HMMER | -0.031600809 | -0.077547608 | 0.01434599 | 0.529395926 |
| dbCAN_4 | dbCAN_3:HMMER | 0.030219608 | -0.015727191 | 0.076166408 | 0.603220029 |
| dbCAN_2:DIAMOND | dbCAN_2:HMMER | 0.029895141 | -0.016051658 | 0.075841941 | 0.620411343 |
| CUPP | dbCAN_2:DIAMOND | -0.029865745 | -0.075812544 | 0.016081054 | 0.621963026 |
| dbCAN_4:HMMER | dbCAN_4 | -0.02985729 | -0.075804089 | 0.01608951 | 0.622409135 |
| dbCAN_4:dbCANsub | dbCAN_3:DIAMOND | -0.028759953 | -0.074706753 | 0.017186846 | 0.679322585 |
| dbCAN_4:dbCANsub | dbCAN_3:HMMER | 0.027446835 | -0.018499964 | 0.073393635 | 0.743637113 |
| dbCAN_2:HMMER | dbCAN_2 | -0.027402448 | -0.073349248 | 0.018544351 | 0.745716545 |
| CUPP | dbCAN_2 | -0.027373052 | -0.073319851 | 0.018573747 | 0.74708988 |
| dbCAN_4:dbCANsub | dbCAN_4:HMMER | 0.027084517 | -0.018862283 | 0.073031316 | 0.760403427 |
| dbCAN_4:dbCANsub | dbCAN_4:DIAMOND | -0.026409765 | -0.072356564 | 0.019537034 | 0.790281547 |
| dbCAN_4 | dbCAN_3:DIAMOND | -0.025987181 | -0.07193398 | 0.019959619 | 0.80802543 |
| dbCAN_4:DIAMOND | dbCAN_4 | 0.023636992 | -0.022309807 | 0.069583791 | 0.891309516 |
| dbCAN_3 | dbCAN_2 | 0.021345986 | -0.024600813 | 0.067292785 | 0.946020509 |
| dbCAN_3:DIAMOND | dbCAN_3 | 0.019297606 | -0.026649193 | 0.065244406 | 0.975079473 |
| dbCAN_3 | dbCAN_2:DIAMOND | 0.018853293 | -0.027093506 | 0.064800092 | 0.979366195 |
| dbCAN_3:HMMER | dbCAN_2:DIAMOND | -0.018055889 | -0.064002689 | 0.02789091 | 0.985599472 |
| dbCAN_4:HMMER | dbCAN_2:DIAMOND | -0.017693571 | -0.06364037 | 0.028253229 | 0.987884334 |
| dbCAN_4:DIAMOND | dbCAN_3 | 0.016947418 | -0.028999382 | 0.062894217 | 0.991676279 |
| dbCAN_3:HMMER | dbCAN_2 | -0.015563196 | -0.061509996 | 0.030383603 | 0.996156392 |
| dbCAN_4:HMMER | dbCAN_2 | -0.015200878 | -0.061147677 | 0.030745922 | 0.996915968 |
| dbCAN_4 | dbCAN_2 | 0.014656412 | -0.031290387 | 0.060603211 | 0.997818245 |
| dbCAN_3:eCAMI | dbCAN_2:Hotpep | 0.012275634 | -0.033671165 | 0.058222433 | 0.999624875 |
| dbCAN_4:HMMER | dbCAN_2:HMMER | 0.012201571 | -0.033745228 | 0.05814837 | 0.999647474 |
| CUPP | dbCAN_4:HMMER | -0.012172174 | -0.058118974 | 0.033774625 | 0.999656109 |
| dbCAN_4 | dbCAN_2:DIAMOND | 0.012163719 | -0.03378308 | 0.058110518 | 0.999658559 |
| dbCAN_4:dbCANsub | dbCAN_2 | 0.011883639 | -0.03406316 | 0.057830438 | 0.999731551 |
| dbCAN_3:HMMER | dbCAN_2:HMMER | 0.011839252 | -0.034107547 | 0.057786051 | 0.999741766 |
| CUPP | dbCAN_3:HMMER | -0.011809855 | -0.057756655 | 0.034136944 | 0.999748343 |
| dbCAN_4:dbCANsub | dbCAN_3 | -0.009462347 | -0.055409146 | 0.036484452 | 0.999976402 |
| dbCAN_4:dbCANsub | dbCAN_2:DIAMOND | 0.009390946 | -0.036555853 | 0.055337745 | 0.99997828 |
| dbCAN_4 | dbCAN_3 | -0.006689574 | -0.052636373 | 0.039257225 | 0.99999952 |
| dbCAN_4:dbCANsub | dbCAN_4 | -0.002772773 | -0.048719572 | 0.043174026 | 1 |
| dbCAN_2:DIAMOND | dbCAN_2 | 0.002492693 | -0.043454106 | 0.048439492 | 1 |
| dbCAN_4:DIAMOND | dbCAN_3:DIAMOND | -0.002350189 | -0.048296988 | 0.043596611 | 1 |
| CUPP | dbCAN_2:HMMER | 2.94E-05 | -0.045917403 | 0.045976196 | 1 |
| dbCAN_4:HMMER | dbCAN_3:HMMER | 0.000362319 | -0.04558448 | 0.046309118 | 1 |

### 6.2 Glycosyltransferases

**SI table 35: Output of Tukey HSD test for statistically significant differences between the mean F1-score of tools classifying GT CAZyme domains (overleaf)**

| Classifier 1 | Classifier 2 | Mean Difference | Lower 95% Confidence Interval | Upper 95% Confidence Interval | Adjusted P-value |
|---|---|---|---|---|---|
| dbCAN_3:DIAMOND | dbCAN_2:Hotpep | 0.155472051 | 0.107482702 | 0.203461399 | 3.95E-13 |
| dbCAN_2:Hotpep | dbCAN_2:DIAMOND | -0.125400816 | -0.173390165 | -0.077411468 | 4.00E-13 |
| dbCAN_4 | dbCAN_2:Hotpep | 0.146856211 | 0.098866863 | 0.19484556 | 4.06E-13 |
| dbCAN_4:DIAMOND | dbCAN_2:Hotpep | 0.154151682 | 0.106162334 | 0.20214103 | 4.07E-13 |
| dbCAN_3 | dbCAN_2:Hotpep | 0.139703344 | 0.091713995 | 0.187692692 | 4.08E-13 |
| dbCAN_4:dbCANsub | dbCAN_2:Hotpep | 0.145487158 | 0.09749781 | 0.193476507 | 4.08E-13 |
| dbCAN_2:Hotpep | dbCAN_2 | -0.104929049 | -0.152918397 | -0.0569397 | 5.96E-11 |
| dbCAN_4:HMMER | dbCAN_2:Hotpep | 0.096156486 | 0.048167138 | 0.144145834 | 3.62E-09 |
| dbCAN_3:HMMER | dbCAN_2:Hotpep | 0.095959397 | 0.047970048 | 0.143948745 | 3.95E-09 |
| dbCAN_2:Hotpep | dbCAN_2:HMMER | -0.094293511 | -0.142282859 | -0.046304163 | 8.32E-09 |
| CUPP | dbCAN_2:Hotpep | 0.08982508 | 0.041835732 | 0.137814429 | 5.77E-08 |
| dbCAN_3:eCAMI | dbCAN_2:Hotpep | 0.083677867 | 0.035688519 | 0.131667216 | 7.23E-07 |
| dbCAN_3:eCAMI | dbCAN_3:DIAMOND | -0.071794183 | -0.119783532 | -0.023804835 | 5.97E-05 |
| dbCAN_4:DIAMOND | dbCAN_3:eCAMI | 0.070473814 | 0.022484466 | 0.118463163 | 9.37E-05 |
| CUPP | dbCAN_3:DIAMOND | -0.06564697 | -0.113636319 | -0.017657622 | 0.00045302 |
| CUPP | dbCAN_4:DIAMOND | -0.064326602 | -0.11231595 | -0.016337253 | 0.000683255 |
| dbCAN_4 | dbCAN_3:eCAMI | 0.063178344 | 0.015188996 | 0.111167692 | 0.000969738 |
| dbCAN_4:dbCANsub | dbCAN_3:eCAMI | 0.061809291 | 0.013819943 | 0.109798639 | 0.001459144 |
| dbCAN_3:DIAMOND | dbCAN_2:HMMER | 0.06117854 | 0.013189191 | 0.109167888 | 0.001755553 |
| dbCAN_4:DIAMOND | dbCAN_2:HMMER | 0.059858171 | 0.011868823 | 0.107847519 | 0.00256781 |
| dbCAN_3:DIAMOND | dbCAN_3:HMMER | 0.059512654 | 0.011523306 | 0.107502002 | 0.002832087 |
| dbCAN_4:HMMER | dbCAN_3:DIAMOND | -0.059315565 | -0.107304913 | -0.011326216 | 0.002993975 |
| dbCAN_4:DIAMOND | dbCAN_3:HMMER | 0.058192285 | 0.010202937 | 0.106181633 | 0.004093309 |
| dbCAN_4:DIAMOND | dbCAN_4:HMMER | 0.057995196 | 0.010005847 | 0.105984544 | 0.004321101 |
| CUPP | dbCAN_4 | -0.057031131 | -0.105020479 | -0.009041783 | 0.005614147 |
| dbCAN_3:eCAMI | dbCAN_3 | -0.056025476 | -0.104014824 | -0.008036128 | 0.007335766 |
| CUPP | dbCAN_4:dbCANsub | -0.055662078 | -0.103651427 | -0.00767273 | 0.008068674 |
| dbCAN_4 | dbCAN_2:HMMER | 0.0525627 | 0.004573352 | 0.100552049 | 0.017612733 |
| dbCAN_4:dbCANsub | dbCAN_2:HMMER | 0.051193648 | 0.003204299 | 0.099182996 | 0.024408272 |
| dbCAN_4 | dbCAN_3:HMMER | 0.050896815 | 0.002907466 | 0.098886163 | 0.026157258 |
| dbCAN_4:HMMER | dbCAN_4 | -0.050699725 | -0.098689074 | -0.002710377 | 0.027378811 |
| dbCAN_3:DIAMOND | dbCAN_2 | 0.050543002 | 0.002553654 | 0.09853235 | 0.028385772 |
| CUPP | dbCAN_3 | -0.049878263 | -0.097867612 | -0.001888915 | 0.03302753 |
| dbCAN_4:dbCANsub | dbCAN_3:HMMER | 0.049527762 | 0.001538413 | 0.09751711 | 0.035732273 |
| dbCAN_4:dbCANsub | dbCAN_4:HMMER | 0.049330672 | 0.001341324 | 0.097320021 | 0.037336256 |
| dbCAN_4:DIAMOND | dbCAN_2 | 0.049222633 | 0.001233285 | 0.097211982 | 0.038241756 |
| dbCAN_3 | dbCAN_2:HMMER | 0.045409833 | -0.002579516 | 0.093399181 | 0.084729713 |
| dbCAN_3:HMMER | dbCAN_3 | -0.043743947 | -0.091733295 | 0.004245401 | 0.116187017 |
| dbCAN_4:HMMER | dbCAN_3 | -0.043546857 | -0.091536206 | 0.004442491 | 0.120448398 |
| dbCAN_4 | dbCAN_2 | 0.041927163 | -0.006062186 | 0.089916511 | 0.160177302 |
| dbCAN_3:eCAMI | dbCAN_2:DIAMOND | -0.041722949 | -0.089712297 | 0.0062664 | 0.165805519 |
| dbCAN_4:dbCANsub | dbCAN_2 | 0.04055811 | -0.007431238 | 0.088547458 | 0.200668064 |
| CUPP | dbCAN_2:DIAMOND | -0.035575736 | -0.083565084 | 0.012413612 | 0.401471321 |
| dbCAN_3 | dbCAN_2 | 0.034774295 | -0.013215053 | 0.082763643 | 0.440287408 |
| dbCAN_2:DIAMOND | dbCAN_2:HMMER | 0.031107305 | -0.016882043 | 0.079096654 | 0.626309035 |
| dbCAN_3:DIAMOND | dbCAN_2:DIAMOND | 0.030071234 | -0.017918114 | 0.078060583 | 0.677728427 |
| dbCAN_3:HMMER | dbCAN_2:DIAMOND | -0.029441419 | -0.077430768 | 0.018547929 | 0.707937647 |
| dbCAN_4:HMMER | dbCAN_2:DIAMOND | -0.02924433 | -0.077233678 | 0.018745018 | 0.717186441 |
| dbCAN_4:DIAMOND | dbCAN_2:DIAMOND | 0.028750866 | -0.019238483 | 0.076740214 | 0.739856845 |
| dbCAN_4 | dbCAN_2:DIAMOND | 0.021455395 | -0.026533953 | 0.069444743 | 0.959414502 |
| dbCAN_3:eCAMI | dbCAN_2 | -0.021251181 | -0.069240529 | 0.026738167 | 0.962264008 |
| dbCAN_2:DIAMOND | dbCAN_2 | 0.020471768 | -0.027517581 | 0.068461116 | 0.97178651 |
| dbCAN_4:dbCANsub | dbCAN_2:DIAMOND | 0.020086342 | -0.027903006 | 0.068075691 | 0.975758151 |
| dbCAN_3:DIAMOND | dbCAN_3 | 0.015768707 | -0.032220641 | 0.063758055 | 0.997107885 |
| CUPP | dbCAN_2 | -0.015103968 | -0.063093317 | 0.03288538 | 0.99808194 |
| dbCAN_4:DIAMOND | dbCAN_3 | 0.014448338 | -0.03354101 | 0.062437687 | 0.998754337 |
| dbCAN_3 | dbCAN_2:DIAMOND | 0.014302527 | -0.033686821 | 0.062291876 | 0.998872656 |
| dbCAN_4:HMMER | dbCAN_3:eCAMI | 0.012478619 | -0.03551073 | 0.060467967 | 0.999716235 |
| dbCAN_3:eCAMI | dbCAN_3:HMMER | -0.012281529 | -0.060270878 | 0.035707819 | 0.999759418 |
| dbCAN_2:HMMER | dbCAN_2 | -0.010635538 | -0.058624886 | 0.037353811 | 0.999947629 |
| dbCAN_3:eCAMI | dbCAN_2:HMMER | -0.010615643 | -0.058604992 | 0.037373705 | 0.999948675 |
| dbCAN_4:dbCANsub | dbCAN_3:DIAMOND | -0.009984892 | -0.057974241 | 0.038004456 | 0.999973602 |
| dbCAN_3:HMMER | dbCAN_2 | -0.008969652 | -0.056959 | 0.039019697 | 0.999991907 |
| dbCAN_4:HMMER | dbCAN_2 | -0.008772562 | -0.056761911 | 0.039216786 | 0.999993681 |
| dbCAN_4:dbCANsub | dbCAN_4:DIAMOND | -0.008664523 | -0.056653872 | 0.039324825 | 0.999994498 |
| dbCAN_4 | dbCAN_3:DIAMOND | -0.008615839 | -0.056605188 | 0.039373509 | 0.999994834 |
| dbCAN_4:DIAMOND | dbCAN_4 | 0.007295471 | -0.040693878 | 0.055284819 | 0.999999213 |
| dbCAN_4 | dbCAN_3 | 0.007152868 | -0.040836481 | 0.055142216 | 0.999999372 |
| CUPP | dbCAN_4:HMMER | -0.006331406 | -0.054320754 | 0.041657942 | 0.999999845 |
| CUPP | dbCAN_3:eCAMI | 0.006147213 | -0.041842136 | 0.054136561 | 0.99999989 |
| CUPP | dbCAN_3:HMMER | -0.006134317 | -0.054123665 | 0.041855032 | 0.999999893 |
| dbCAN_4:dbCANsub | dbCAN_3 | 0.005783815 | -0.042205533 | 0.053773163 | 0.999999946 |
| CUPP | dbCAN_2:HMMER | -0.004468431 | -0.052457779 | 0.043520918 | 0.999999997 |
| dbCAN_3:HMMER | dbCAN_2:HMMER | 0.001665886 | -0.046323462 | 0.049655234 | 1 |
| dbCAN_4:HMMER | dbCAN_2:HMMER | 0.001862975 | -0.046126373 | 0.049852324 | 1 |
| dbCAN_4:HMMER | dbCAN_3:HMMER | 0.000197089 | -0.047792259 | 0.048186438 | 1 |
| dbCAN_4:DIAMOND | dbCAN_3:DIAMOND | -0.001320369 | -0.049309717 | 0.04666898 | 1 |
| dbCAN_4:dbCANsub | dbCAN_4 | -0.001369053 | -0.049358401 | 0.046620295 | 1 |

**SI table 36: Output of Tukey HSD test for statistically significant differences between the mean accuracy of tools classifying GT CAZyme domains (overleaf)**

| Classifier 1 | Classifier 2 | Mean Difference | Lower 95% Confidence Interval | Upper 95% Confidence Interval | Adjusted P-value |
|---|---|---|---|---|---|
| dbCAN_3:DIAMOND | dbCAN_2:Hotpep | 0.08170861 | 0.046745857 | 0.116671364 | 1.98E-12 |
| dbCAN_4:DIAMOND | dbCAN_2:Hotpep | 0.080718196 | 0.045755442 | 0.115680949 | 3.58E-12 |
| dbCAN_4 | dbCAN_2:Hotpep | 0.075054898 | 0.040092145 | 0.110017652 | 1.49E-10 |
| dbCAN_4:dbCANsub | dbCAN_2:Hotpep | 0.074221285 | 0.039258532 | 0.109184039 | 2.57E-10 |
| dbCAN_3 | dbCAN_2:Hotpep | 0.07203231 | 0.037069556 | 0.106995063 | 1.05E-09 |
| dbCAN_2:Hotpep | dbCAN_2:DIAMOND | -0.064074443 | -0.099037197 | -0.02911169 | 1.27E-07 |
| dbCAN_2:Hotpep | dbCAN_2 | -0.051940721 | -0.086903475 | -0.016977968 | 7.09E-05 |
| dbCAN_3:HMMER | dbCAN_2:Hotpep | 0.047189064 | 0.01222631 | 0.082151817 | 0.000595585 |
| dbCAN_4:HMMER | dbCAN_2:Hotpep | 0.047185494 | 0.01222274 | 0.082148247 | 0.000596491 |
| dbCAN_2:Hotpep | dbCAN_2:HMMER | -0.046714359 | -0.081677112 | -0.011751605 | 0.000728178 |
| dbCAN_3:eCAMI | dbCAN_3:DIAMOND | -0.043524251 | -0.078487005 | -0.008561498 | 0.002654941 |
| CUPP | dbCAN_2:Hotpep | 0.043124154 | 0.0081614 | 0.078086907 | 0.003100202 |
| dbCAN_4:DIAMOND | dbCAN_3:eCAMI | 0.042533837 | 0.007571083 | 0.07749659 | 0.003885348 |
| CUPP | dbCAN_3:DIAMOND | -0.038584457 | -0.07354721 | -0.003621703 | 0.015985161 |
| dbCAN_3:eCAMI | dbCAN_2:Hotpep | 0.038184359 | 0.003221606 | 0.073147113 | 0.0182704 |
| CUPP | dbCAN_4:DIAMOND | -0.037594042 | -0.072556796 | -0.002631289 | 0.022176797 |
| dbCAN_4 | dbCAN_3:eCAMI | 0.036870539 | 0.001907786 | 0.071833293 | 0.027964612 |
| dbCAN_4:dbCANsub | dbCAN_3:eCAMI | 0.036036926 | 0.001074173 | 0.07099968 | 0.03624659 |
| dbCAN_3:DIAMOND | dbCAN_2:HMMER | 0.034994251 | 3.15E-05 | 0.069957005 | 0.04954076 |
| dbCAN_4:HMMER | dbCAN_3:DIAMOND | -0.034523117 | -0.06948587 | 0.000439637 | 0.056797937 |
| dbCAN_3:DIAMOND | dbCAN_3:HMMER | 0.034519546 | -0.000443207 | 0.0694823 | 0.05685619 |
| dbCAN_4:DIAMOND | dbCAN_2:HMMER | 0.034003837 | -0.000958917 | 0.06896659 | 0.065817234 |
| dbCAN_3:eCAMI | dbCAN_3 | -0.033847951 | -0.068810704 | 0.001114803 | 0.068747745 |
| dbCAN_4:DIAMOND | dbCAN_4:HMMER | 0.033532702 | -0.001430052 | 0.068495455 | 0.075006088 |
| dbCAN_4:DIAMOND | dbCAN_3:HMMER | 0.033529132 | -0.001433622 | 0.068491885 | 0.075079569 |
| CUPP | dbCAN_4 | -0.031930745 | -0.066893498 | 0.003032009 | 0.114413083 |
| CUPP | dbCAN_4:dbCANsub | -0.031097132 | -0.066059885 | 0.003865622 | 0.140558906 |
| dbCAN_3:DIAMOND | dbCAN_2 | 0.029767889 | -0.005194865 | 0.064730642 | 0.191210206 |
| CUPP | dbCAN_3 | -0.028908156 | -0.06387091 | 0.006054597 | 0.23013441 |
| dbCAN_4:DIAMOND | dbCAN_2 | 0.028777474 | -0.006185279 | 0.063740228 | 0.236479748 |
| dbCAN_4 | dbCAN_2:HMMER | 0.028340539 | -0.006622214 | 0.063303293 | 0.258508744 |
| dbCAN_4:HMMER | dbCAN_4 | -0.027869405 | -0.062832158 | 0.007093349 | 0.283641316 |
| dbCAN_4 | dbCAN_3:HMMER | 0.027865834 | -0.007096919 | 0.062828588 | 0.283837113 |
| dbCAN_4:dbCANsub | dbCAN_2:HMMER | 0.027506927 | -0.007455827 | 0.06246968 | 0.303921777 |
| dbCAN_4:dbCANsub | dbCAN_4:HMMER | 0.027035792 | -0.007926962 | 0.061998545 | 0.331452335 |
| dbCAN_4:dbCANsub | dbCAN_3:HMMER | 0.027032222 | -0.007930532 | 0.061994975 | 0.331665811 |
| dbCAN_3:eCAMI | dbCAN_2:DIAMOND | -0.025890084 | -0.060852838 | 0.009072669 | 0.403348185 |
| dbCAN_3 | dbCAN_2:HMMER | 0.025317951 | -0.009644802 | 0.060280705 | 0.441432208 |
| dbCAN_4:HMMER | dbCAN_3 | -0.024846816 | -0.05980957 | 0.010115937 | 0.473609107 |
| dbCAN_3:HMMER | dbCAN_3 | -0.024843246 | -0.059806 | 0.010119507 | 0.473855222 |
| dbCAN_4 | dbCAN_2 | 0.023114177 | -0.011848577 | 0.05807693 | 0.594907457 |
| dbCAN_4:dbCANsub | dbCAN_2 | 0.022280564 | -0.012682189 | 0.057243318 | 0.652605946 |
| CUPP | dbCAN_2:DIAMOND | -0.02095029 | -0.055913043 | 0.014012464 | 0.739622415 |
| dbCAN_3 | dbCAN_2 | 0.020091588 | -0.014871165 | 0.055054342 | 0.790543953 |
| dbCAN_3:DIAMOND | dbCAN_2:DIAMOND | 0.017634167 | -0.017328587 | 0.05259692 | 0.904453849 |
| dbCAN_2:DIAMOND | dbCAN_2:HMMER | 0.017360084 | -0.017602669 | 0.052322838 | 0.913935328 |
| dbCAN_4:HMMER | dbCAN_2:DIAMOND | -0.01688895 | -0.051851703 | 0.018073804 | 0.928727249 |
| dbCAN_3:HMMER | dbCAN_2:DIAMOND | -0.01688538 | -0.051848133 | 0.018077374 | 0.928832184 |
| dbCAN_4:DIAMOND | dbCAN_2:DIAMOND | 0.016643752 | -0.018319001 | 0.051606506 | 0.935689219 |
| dbCAN_3:eCAMI | dbCAN_2 | -0.013756362 | -0.048719116 | 0.021206391 | 0.985448854 |
| dbCAN_2:DIAMOND | dbCAN_2 | 0.012133722 | -0.022829032 | 0.047096475 | 0.995191027 |
| dbCAN_4 | dbCAN_2:DIAMOND | 0.010980455 | -0.023982299 | 0.045943208 | 0.998121177 |
| dbCAN_4:dbCANsub | dbCAN_2:DIAMOND | 0.010146842 | -0.024815911 | 0.045109596 | 0.99913375 |
| dbCAN_3:DIAMOND | dbCAN_3 | 0.0096763 | -0.025286453 | 0.044639054 | 0.999462523 |
| dbCAN_3:eCAMI | dbCAN_3:HMMER | -0.009004705 | -0.043967458 | 0.025958049 | 0.999743029 |
| dbCAN_4:HMMER | dbCAN_3:eCAMI | 0.009001135 | -0.025961619 | 0.043963888 | 0.999744084 |
| CUPP | dbCAN_2 | -0.008816568 | -0.043779321 | 0.026146186 | 0.999793721 |
| dbCAN_4:DIAMOND | dbCAN_3 | 0.008685886 | -0.026276868 | 0.043648639 | 0.999823553 |
| dbCAN_3:eCAMI | dbCAN_2:HMMER | -0.00853 | -0.043492753 | 0.026432754 | 0.99985413 |
| dbCAN_3 | dbCAN_2:DIAMOND | 0.007957867 | -0.027004887 | 0.04292062 | 0.999930253 |
| dbCAN_4:dbCANsub | dbCAN_3:DIAMOND | -0.007487325 | -0.042450078 | 0.027475429 | 0.999963864 |
| dbCAN_4 | dbCAN_3:DIAMOND | -0.006653712 | -0.041616465 | 0.028309042 | 0.999990113 |
| dbCAN_4:dbCANsub | dbCAN_4:DIAMOND | -0.00649691 | -0.041459664 | 0.028465843 | 0.999992415 |
| dbCAN_4:DIAMOND | dbCAN_4 | 0.005663297 | -0.029299456 | 0.040626051 | 0.999998379 |
| dbCAN_2:HMMER | dbCAN_2 | -0.005226363 | -0.040189116 | 0.029736391 | 0.99999935 |
| CUPP | dbCAN_3:eCAMI | 0.004939794 | -0.030022959 | 0.039902548 | 0.99999966 |
| dbCAN_4:HMMER | dbCAN_2 | -0.004755228 | -0.039717981 | 0.030207526 | 0.99999978 |
| dbCAN_3:HMMER | dbCAN_2 | -0.004751658 | -0.039714411 | 0.030211096 | 0.999999782 |
| CUPP | dbCAN_3:HMMER | -0.00406491 | -0.039027664 | 0.030897843 | 0.999999964 |
| CUPP | dbCAN_4:HMMER | -0.00406134 | -0.039024094 | 0.030901413 | 0.999999965 |
| CUPP | dbCAN_2:HMMER | -0.003590205 | -0.038552959 | 0.031372548 | 0.999999992 |
| dbCAN_4 | dbCAN_3 | 0.003022588 | -0.031940165 | 0.037985342 | 0.999999999 |
| dbCAN_4:dbCANsub | dbCAN_3 | 0.002188976 | -0.032773778 | 0.037151729 | 1 |
| dbCAN_3:HMMER | dbCAN_2:HMMER | 0.000474705 | -0.034488049 | 0.035437458 | 1 |
| dbCAN_4:HMMER | dbCAN_2:HMMER | 0.000471135 | -0.034491619 | 0.035433888 | 1 |
| dbCAN_4:HMMER | dbCAN_3:HMMER | -3.57E-06 | -0.034966324 | 0.034959183 | 1 |
| dbCAN_4:DIAMOND | dbCAN_3:DIAMOND | -0.000990415 | -0.035953168 | 0.033972339 | 1 |
| dbCAN_4:dbCANsub | dbCAN_4 | -0.000833613 | -0.035796366 | 0.034129141 | 1 |

**SI table 37: Output of Tukey HSD test for statistically significant differences between the mean sensitivity of tools classifying GT CAZyme domains (overleaf)**

| Classifier 1 | Classifier 2 | Mean Difference | Lower 95% Confidence Interval | Upper 95% Confidence Interval | Adjusted P-value |
|---|---|---|---|---|---|
| dbCAN_3:DIAMOND | dbCAN_2:Hotpep | 0.251972385 | 0.18873865 | 0.31520612 | 2.86E-13 |
| dbCAN_4 | dbCAN_2:Hotpep | 0.232412765 | 0.16917903 | 0.295646501 | 2.86E-13 |
| dbCAN_4:DIAMOND | dbCAN_2:Hotpep | 0.249687312 | 0.186453576 | 0.312921047 | 2.86E-13 |
| dbCAN_4:dbCANsub | dbCAN_2:Hotpep | 0.22840387 | 0.165170134 | 0.291637605 | 2.86E-13 |
| dbCAN_3 | dbCAN_2:Hotpep | 0.216703259 | 0.153469524 | 0.279936994 | 3.06E-13 |
| dbCAN_2:Hotpep | dbCAN_2:DIAMOND | -0.200103226 | -0.263336961 | -0.136869491 | 4.05E-13 |
| dbCAN_2:Hotpep | dbCAN_2 | -0.159091954 | -0.222325689 | -0.095858219 | 4.39E-13 |
| dbCAN_4:HMMER | dbCAN_2:Hotpep | 0.140284156 | 0.077050421 | 0.203517892 | 2.83E-11 |
| dbCAN_3:HMMER | dbCAN_2:Hotpep | 0.13997216 | 0.076738424 | 0.203205895 | 3.18E-11 |
| dbCAN_2:Hotpep | dbCAN_2:HMMER | -0.137265788 | -0.200499523 | -0.074032053 | 8.59E-11 |
| CUPP | dbCAN_2:Hotpep | 0.128159899 | 0.064926164 | 0.191393635 | 2.19E-09 |
| dbCAN_3:eCAMI | dbCAN_3:DIAMOND | -0.127365443 | -0.190599178 | -0.064131708 | 2.88E-09 |
| dbCAN_4:DIAMOND | dbCAN_3:eCAMI | 0.12508037 | 0.061846635 | 0.188314105 | 6.28E-09 |
| dbCAN_3:eCAMI | dbCAN_2:Hotpep | 0.124606942 | 0.061373206 | 0.187840677 | 7.37E-09 |
| CUPP | dbCAN_3:DIAMOND | -0.123812485 | -0.187046221 | -0.06057875 | 9.62E-09 |
| CUPP | dbCAN_4:DIAMOND | -0.121527412 | -0.184761147 | -0.058293677 | 2.06E-08 |
| dbCAN_3:DIAMOND | dbCAN_2:HMMER | 0.114706597 | 0.051472862 | 0.177940332 | 1.84E-07 |
| dbCAN_4:DIAMOND | dbCAN_2:HMMER | 0.112421524 | 0.049187788 | 0.175655259 | 3.74E-07 |
| dbCAN_3:DIAMOND | dbCAN_3:HMMER | 0.112000225 | 0.04876649 | 0.17523396 | 4.26E-07 |
| dbCAN_4:HMMER | dbCAN_3:DIAMOND | -0.111688228 | -0.174921964 | -0.048454493 | 4.68E-07 |
| dbCAN_4:DIAMOND | dbCAN_3:HMMER | 0.109715152 | 0.046481417 | 0.172948887 | 8.52E-07 |
| dbCAN_4:DIAMOND | dbCAN_4:HMMER | 0.109403155 | 0.04616942 | 0.17263689 | 9.35E-07 |
| dbCAN_4 | dbCAN_3:eCAMI | 0.107805824 | 0.044572088 | 0.171039559 | 1.50E-06 |
| CUPP | dbCAN_4 | -0.104252866 | -0.167486601 | -0.041019131 | 4.24E-06 |
| dbCAN_4:dbCANsub | dbCAN_3:eCAMI | 0.103796928 | 0.040563193 | 0.167030663 | 4.83E-06 |
| CUPP | dbCAN_4:dbCANsub | -0.10024397 | -0.163477706 | -0.037010235 | 1.31E-05 |
| dbCAN_4 | dbCAN_2:HMMER | 0.095146977 | 0.031913242 | 0.158380713 | 5.18E-05 |
| dbCAN_3:DIAMOND | dbCAN_2 | 0.092880431 | 0.029646696 | 0.156114166 | 9.32E-05 |
| dbCAN_4 | dbCAN_3:HMMER | 0.092440606 | 0.02920687 | 0.155674341 | 0.000104362 |
| dbCAN_4:HMMER | dbCAN_4 | -0.092128609 | -0.155362344 | -0.028894874 | 0.000113004 |
| dbCAN_3:eCAMI | dbCAN_3 | -0.092096317 | -0.155330052 | -0.028862582 | 0.000113937 |
| dbCAN_4:dbCANsub | dbCAN_2:HMMER | 0.091138082 | 0.027904347 | 0.154371817 | 0.000145219 |
| dbCAN_4:DIAMOND | dbCAN_2 | 0.090595358 | 0.027361622 | 0.153829093 | 0.000166419 |
| CUPP | dbCAN_3 | -0.088543359 | -0.151777095 | -0.025309624 | 0.00027652 |
| dbCAN_4:dbCANsub | dbCAN_3:HMMER | 0.08843171 | 0.025197975 | 0.151665445 | 0.00028417 |
| dbCAN_4:dbCANsub | dbCAN_4:HMMER | 0.088119713 | 0.024885978 | 0.151353449 | 0.000306629 |
| dbCAN_3 | dbCAN_2:HMMER | 0.079437471 | 0.016203736 | 0.142671206 | 0.002272009 |
| dbCAN_3:HMMER | dbCAN_3 | -0.076731099 | -0.139964835 | -0.013497364 | 0.004047919 |
| dbCAN_4:HMMER | dbCAN_3 | -0.076419103 | -0.139652838 | -0.013185367 | 0.004320163 |
| dbCAN_3:eCAMI | dbCAN_2:DIAMOND | -0.075496284 | -0.13873002 | -0.012262549 | 0.005227769 |
| dbCAN_4 | dbCAN_2 | 0.073320811 | 0.010087076 | 0.136554546 | 0.00810549 |
| CUPP | dbCAN_2:DIAMOND | -0.071943327 | -0.135177062 | -0.008709591 | 0.010613242 |
| dbCAN_4:dbCANsub | dbCAN_2 | 0.069311916 | 0.00607818 | 0.132545651 | 0.0174447 |
| dbCAN_2:DIAMOND | dbCAN_2:HMMER | 0.062837438 | -0.000396297 | 0.126071173 | 0.053296226 |
| dbCAN_3:HMMER | dbCAN_2:DIAMOND | -0.060131066 | -0.123364802 | 0.003102669 | 0.081071483 |
| dbCAN_4:HMMER | dbCAN_2:DIAMOND | -0.05981907 | -0.123052805 | 0.003414666 | 0.084927727 |
| dbCAN_3 | dbCAN_2 | 0.057611305 | -0.00562243 | 0.12084504 | 0.116647649 |
| dbCAN_3:DIAMOND | dbCAN_2:DIAMOND | 0.051869159 | -0.011364576 | 0.115102894 | 0.241328995 |
| dbCAN_4:DIAMOND | dbCAN_2:DIAMOND | 0.049584086 | -0.01364965 | 0.112817821 | 0.309152123 |
| dbCAN_2:DIAMOND | dbCAN_2 | 0.041011272 | -0.022222463 | 0.104245007 | 0.625453236 |
| dbCAN_3:DIAMOND | dbCAN_3 | 0.035269126 | -0.027964609 | 0.098502861 | 0.822512347 |
| dbCAN_3:eCAMI | dbCAN_2 | -0.034485012 | -0.097718748 | 0.028748723 | 0.844202186 |
| dbCAN_4:DIAMOND | dbCAN_3 | 0.032984053 | -0.030249683 | 0.096217788 | 0.881278014 |
| dbCAN_4 | dbCAN_2:DIAMOND | 0.032309539 | -0.030924196 | 0.095543275 | 0.895984837 |
| CUPP | dbCAN_2 | -0.030932055 | -0.09416579 | 0.032301681 | 0.92224865 |
| dbCAN_4:dbCANsub | dbCAN_2:DIAMOND | 0.028300644 | -0.034933092 | 0.091534379 | 0.959090666 |
| dbCAN_4:dbCANsub | dbCAN_3:DIAMOND | -0.023568515 | -0.08680225 | 0.03966522 | 0.990874399 |
| dbCAN_2:HMMER | dbCAN_2 | -0.021826166 | -0.085059901 | 0.041407569 | 0.995425036 |
| dbCAN_4:dbCANsub | dbCAN_4:DIAMOND | -0.021283442 | -0.084517177 | 0.041950293 | 0.996376355 |
| dbCAN_4 | dbCAN_3:DIAMOND | -0.01955962 | -0.082793355 | 0.043674116 | 0.998378164 |
| dbCAN_3:HMMER | dbCAN_2 | -0.019119794 | -0.08235353 | 0.044113941 | 0.998700845 |
| dbCAN_4:HMMER | dbCAN_2 | -0.018807798 | -0.082041533 | 0.044425938 | 0.998894936 |
| dbCAN_4:DIAMOND | dbCAN_4 | 0.017274546 | -0.045959189 | 0.080508282 | 0.999529026 |
| dbCAN_3 | dbCAN_2:DIAMOND | 0.016600033 | -0.046633702 | 0.079833768 | 0.999686883 |
| dbCAN_4 | dbCAN_3 | 0.015709506 | -0.047524229 | 0.078943242 | 0.999823531 |
| dbCAN_4:HMMER | dbCAN_3:eCAMI | 0.015677215 | -0.047556521 | 0.07891095 | 0.999827298 |
| dbCAN_3:eCAMI | dbCAN_3:HMMER | -0.015365218 | -0.078598953 | 0.047868517 | 0.999860208 |
| dbCAN_3:eCAMI | dbCAN_2:HMMER | -0.012658846 | -0.075892581 | 0.050574889 | 0.999982704 |
| CUPP | dbCAN_4:HMMER | -0.012124257 | -0.075357992 | 0.051109478 | 0.999989258 |
| CUPP | dbCAN_3:HMMER | -0.01181226 | -0.075045995 | 0.051421475 | 0.999991958 |
| dbCAN_4:dbCANsub | dbCAN_3 | 0.011700611 | -0.051533124 | 0.074934346 | 0.999992764 |
| CUPP | dbCAN_2:HMMER | -0.009105888 | -0.072339624 | 0.054127847 | 0.999999577 |
| dbCAN_4:dbCANsub | dbCAN_4 | -0.004008896 | -0.067242631 | 0.05922484 | 1 |
| CUPP | dbCAN_3:eCAMI | 0.003552958 | -0.059680777 | 0.066786693 | 1 |
| dbCAN_4:HMMER | dbCAN_2:HMMER | 0.003018369 | -0.060215367 | 0.066252104 | 1 |
| dbCAN_3:HMMER | dbCAN_2:HMMER | 0.002706372 | -0.060527363 | 0.065940107 | 1 |
| dbCAN_4:HMMER | dbCAN_3:HMMER | 0.000311997 | -0.062921738 | 0.063545732 | 1 |
| dbCAN_4:DIAMOND | dbCAN_3:DIAMOND | -0.002285073 | -0.065518809 | 0.060948662 | 1 |

### 6.3 Polysaccharide lyases

**SI table 38: Output of Tukey HSD test for statistically significant differences between the mean F1-score of tools classifying PL CAZyme domains (overleaf)**

| Classifier 1 | Classifier 2 | Mean Difference | Lower 95% Confidence Interval | Upper 95% Confidence Interval | Adjusted P-value |
|---|---|---|---|---|---|
| dbCAN_3:eCAMI | dbCAN_3 | -0.135329652 | -0.247574462 | -0.023084842 | 0.004559515 |
| dbCAN_4:DIAMOND | dbCAN_3:eCAMI | 0.135190784 | 0.022945974 | 0.247435594 | 0.004633462 |
| dbCAN_3:eCAMI | dbCAN_3:DIAMOND | -0.128098585 | -0.240343395 | -0.015853775 | 0.010268012 |
| dbCAN_4:dbCANsub | dbCAN_3:eCAMI | 0.127463759 | 0.015218949 | 0.239708569 | 0.010998179 |
| dbCAN_3:eCAMI | dbCAN_3:HMMER | -0.126819014 | -0.239063824 | -0.014574204 | 0.011787792 |
| dbCAN_4 | dbCAN_3:eCAMI | 0.126819014 | 0.014574204 | 0.239063824 | 0.011787792 |
| dbCAN_4:HMMER | dbCAN_3:eCAMI | 0.126819014 | 0.014574204 | 0.239063824 | 0.011787792 |
| dbCAN_3 | dbCAN_2:Hotpep | 0.102337662 | -0.010496363 | 0.215171688 | 0.120494331 |
| dbCAN_4:DIAMOND | dbCAN_2:Hotpep | 0.102198794 | -0.010635231 | 0.21503282 | 0.121790043 |
| CUPP | dbCAN_3 | -0.097574325 | -0.21040835 | 0.015259701 | 0.171359841 |
| CUPP | dbCAN_4:DIAMOND | -0.097435457 | -0.210269482 | 0.015398569 | 0.173048336 |
| dbCAN_3:DIAMOND | dbCAN_2:Hotpep | 0.095106596 | -0.01772743 | 0.207940621 | 0.20316678 |
| dbCAN_4:dbCANsub | dbCAN_2:Hotpep | 0.094471769 | -0.018362256 | 0.207305795 | 0.211971332 |
| dbCAN_3:HMMER | dbCAN_2:Hotpep | 0.093827024 | -0.019007001 | 0.20666105 | 0.221175199 |
| dbCAN_4 | dbCAN_2:Hotpep | 0.093827024 | -0.019007001 | 0.20666105 | 0.221175199 |
| dbCAN_4:HMMER | dbCAN_2:Hotpep | 0.093827024 | -0.019007001 | 0.20666105 | 0.221175199 |
| dbCAN_3 | dbCAN_2:DIAMOND | 0.093726909 | -0.019107116 | 0.206560935 | 0.222628002 |
| dbCAN_4:DIAMOND | dbCAN_2:DIAMOND | 0.093588041 | -0.019245984 | 0.206422067 | 0.224653687 |
| CUPP | dbCAN_3:DIAMOND | -0.090343258 | -0.203177284 | 0.022490768 | 0.275428906 |
| CUPP | dbCAN_4:dbCANsub | -0.089708432 | -0.202542457 | 0.023125594 | 0.286119968 |
| CUPP | dbCAN_3:HMMER | -0.089063687 | -0.201897712 | 0.023770339 | 0.297223181 |
| CUPP | dbCAN_4 | -0.089063687 | -0.201897712 | 0.023770339 | 0.297223181 |
| CUPP | dbCAN_4:HMMER | -0.089063687 | -0.201897712 | 0.023770339 | 0.297223181 |
| dbCAN_3:DIAMOND | dbCAN_2:DIAMOND | 0.086495843 | -0.026338183 | 0.199329868 | 0.343784897 |
| dbCAN_4:dbCANsub | dbCAN_2:DIAMOND | 0.085861016 | -0.026973009 | 0.198695042 | 0.355839536 |
| dbCAN_3:HMMER | dbCAN_2:DIAMOND | 0.085216271 | -0.027617754 | 0.198050297 | 0.368287063 |
| dbCAN_4 | dbCAN_2:DIAMOND | 0.085216271 | -0.027617754 | 0.198050297 | 0.368287063 |
| dbCAN_4:HMMER | dbCAN_2:DIAMOND | 0.085216271 | -0.027617754 | 0.198050297 | 0.368287063 |
| dbCAN_3:eCAMI | dbCAN_2:HMMER | -0.077772956 | -0.190017765 | 0.034471854 | 0.514381844 |
| dbCAN_3 | dbCAN_2 | 0.075287193 | -0.037546832 | 0.188121219 | 0.577307729 |
| dbCAN_4:DIAMOND | dbCAN_2 | 0.075148325 | -0.037685701 | 0.18798235 | 0.58032601 |
| dbCAN_3:DIAMOND | dbCAN_2 | 0.068056126 | -0.044777899 | 0.180890152 | 0.728570226 |
| dbCAN_4:dbCANsub | dbCAN_2 | 0.0674213 | -0.045412725 | 0.180255326 | 0.740857412 |
| dbCAN_3:HMMER | dbCAN_2 | 0.066776555 | -0.046057471 | 0.17961058 | 0.75310174 |
| dbCAN_4 | dbCAN_2 | 0.066776555 | -0.046057471 | 0.17961058 | 0.75310174 |
| dbCAN_4:HMMER | dbCAN_2 | 0.066776555 | -0.046057471 | 0.17961058 | 0.75310174 |
| dbCAN_3:eCAMI | dbCAN_2 | -0.060042459 | -0.172287269 | 0.052202351 | 0.859416425 |
| dbCAN_3 | dbCAN_2:HMMER | 0.057556697 | -0.055277329 | 0.170390722 | 0.895599466 |
| dbCAN_4:DIAMOND | dbCAN_2:HMMER | 0.057417828 | -0.055416197 | 0.170251854 | 0.89722192 |
| dbCAN_3:DIAMOND | dbCAN_2:HMMER | 0.05032563 | -0.062508396 | 0.163159655 | 0.95938631 |
| dbCAN_4:dbCANsub | dbCAN_2:HMMER | 0.049690804 | -0.063143222 | 0.162524829 | 0.963126268 |
| dbCAN_3:HMMER | dbCAN_2:HMMER | 0.049046058 | -0.063787967 | 0.161880084 | 0.966656038 |
| dbCAN_4 | dbCAN_2:HMMER | 0.049046058 | -0.063787967 | 0.161880084 | 0.966656038 |
| dbCAN_4:HMMER | dbCAN_2:HMMER | 0.049046058 | -0.063787967 | 0.161880084 | 0.966656038 |
| dbCAN_2:Hotpep | dbCAN_2:HMMER | -0.044780966 | -0.157614991 | 0.06805306 | 0.983994575 |
| dbCAN_3:eCAMI | dbCAN_2:DIAMOND | -0.041602743 | -0.153847553 | 0.070642067 | 0.991098551 |
| CUPP | dbCAN_2:HMMER | -0.040017628 | -0.152851654 | 0.072816397 | 0.993989059 |
| CUPP | dbCAN_3:eCAMI | 0.037755327 | -0.074489483 | 0.150000137 | 0.996300569 |
| dbCAN_2:DIAMOND | dbCAN_2:HMMER | -0.036170213 | -0.149004238 | 0.076663813 | 0.997648506 |
| dbCAN_3:eCAMI | dbCAN_2:Hotpep | -0.03299199 | -0.1452368 | 0.07925282 | 0.99898758 |
| dbCAN_2:Hotpep | dbCAN_2 | -0.027050469 | -0.139884495 | 0.085783556 | 0.99987468 |
| CUPP | dbCAN_2 | -0.022287132 | -0.135121157 | 0.090546894 | 0.999984541 |
| dbCAN_2:DIAMOND | dbCAN_2 | -0.018439716 | -0.131273742 | 0.094394309 | 0.99999814 |
| dbCAN_2:HMMER | dbCAN_2 | 0.017730496 | -0.095103529 | 0.130564522 | 0.999998808 |
| dbCAN_2:Hotpep | dbCAN_2:DIAMOND | -0.008610753 | -0.121444778 | 0.104223273 | 1 |
| CUPP | dbCAN_2:DIAMOND | -0.003847415 | -0.116681441 | 0.10898661 | 1 |
| CUPP | dbCAN_2:Hotpep | 0.004763338 | -0.108070688 | 0.117597363 | 1 |
| dbCAN_3:HMMER | dbCAN_3 | -0.008510638 | -0.121344664 | 0.104323387 | 1 |
| dbCAN_3:DIAMOND | dbCAN_3 | -0.007231067 | -0.120065092 | 0.105602959 | 1 |
| dbCAN_4 | dbCAN_3 | -0.008510638 | -0.121344664 | 0.104323387 | 1 |
| dbCAN_4:HMMER | dbCAN_3 | -0.008510638 | -0.121344664 | 0.104323387 | 1 |
| dbCAN_4:DIAMOND | dbCAN_3 | -0.000138868 | -0.112972894 | 0.112695157 | 1 |
| dbCAN_4:dbCANsub | dbCAN_3 | -0.007865893 | -0.120699919 | 0.104968133 | 1 |
| dbCAN_3:DIAMOND | dbCAN_3:HMMER | 0.001279571 | -0.111554454 | 0.114113597 | 1 |
| dbCAN_4 | dbCAN_3:HMMER | 3.33E-16 | -0.112834026 | 0.112834026 | 1 |
| dbCAN_4:HMMER | dbCAN_3:HMMER | 4.44E-16 | -0.112834026 | 0.112834026 | 1 |
| dbCAN_4:DIAMOND | dbCAN_3:HMMER | 0.00837177 | -0.104462255 | 0.121205796 | 1 |
| dbCAN_4:dbCANsub | dbCAN_3:HMMER | 0.000644745 | -0.11218928 | 0.113478771 | 1 |
| dbCAN_4 | dbCAN_3:DIAMOND | -0.001279571 | -0.114113597 | 0.111554454 | 1 |
| dbCAN_4:HMMER | dbCAN_3:DIAMOND | -0.001279571 | -0.114113597 | 0.111554454 | 1 |
| dbCAN_4:DIAMOND | dbCAN_3:DIAMOND | 0.007092199 | -0.105741827 | 0.119926224 | 1 |
| dbCAN_4:dbCANsub | dbCAN_3:DIAMOND | -0.000634826 | -0.113468852 | 0.112199199 | 1 |
| dbCAN_4:HMMER | dbCAN_4 | 1.11E-16 | -0.112834026 | 0.112834026 | 1 |
| dbCAN_4:DIAMOND | dbCAN_4 | 0.00837177 | -0.104462255 | 0.121205796 | 1 |
| dbCAN_4:dbCANsub | dbCAN_4 | 0.000644745 | -0.11218928 | 0.113478771 | 1 |
| dbCAN_4:DIAMOND | dbCAN_4:HMMER | 0.00837177 | -0.104462255 | 0.121205796 | 1 |
| dbCAN_4:dbCANsub | dbCAN_4:HMMER | 0.000644745 | -0.11218928 | 0.113478771 | 1 |
| dbCAN_4:dbCANsub | dbCAN_4:DIAMOND | -0.007727025 | -0.12056105 | 0.105107001 | 1 |

**SI table 39: Output of Tukey HSD test for statistically significant differences between the mean accuracy of tools classifying PL CAZyme domains (overleaf)**

| Classifier 1 | Classifier 2 | Mean Difference | Lower 95% Confidence Interval | Upper 95% Confidence Interval | Adjusted P-value |
|---|---|---|---|---|---|
| dbCAN_3:eCAMI | dbCAN_3 | -0.006950693 | -0.011818341 | -0.002083045 | 0.000188774 |
| dbCAN_4:dbCANsub | dbCAN_3:eCAMI | 0.006744521 | 0.001876873 | 0.011612169 | 0.000361621 |
| dbCAN_4:DIAMOND | dbCAN_3:eCAMI | 0.006715774 | 0.001848125 | 0.011583422 | 0.000395314 |
| dbCAN_4 | dbCAN_3:eCAMI | 0.006528449 | 0.0016608 | 0.011396097 | 0.00069971 |
| dbCAN_3:eCAMI | dbCAN_3:HMMER | -0.006523096 | -0.011390744 | -0.001655448 | 0.000711046 |
| dbCAN_4:HMMER | dbCAN_3:eCAMI | 0.006523096 | 0.001655448 | 0.011390744 | 0.000711046 |
| dbCAN_3:eCAMI | dbCAN_3:DIAMOND | -0.006502296 | -0.011369944 | -0.001634648 | 0.000756769 |
| dbCAN_3 | dbCAN_2:Hotpep | 0.00607487 | 0.00118167 | 0.010968071 | 0.002838003 |
| dbCAN_4:dbCANsub | dbCAN_2:Hotpep | 0.005868698 | 0.000975498 | 0.010761899 | 0.004948255 |
| dbCAN_4:DIAMOND | dbCAN_2:Hotpep | 0.005839951 | 0.00094675 | 0.010733151 | 0.005337707 |
| dbCAN_4 | dbCAN_2:Hotpep | 0.005652626 | 0.000759425 | 0.010545826 | 0.008651214 |
| dbCAN_3:HMMER | dbCAN_2:Hotpep | 0.005647273 | 0.000754073 | 0.010540473 | 0.008768974 |
| dbCAN_4:HMMER | dbCAN_2:Hotpep | 0.005647273 | 0.000754073 | 0.010540473 | 0.008768974 |
| dbCAN_3:DIAMOND | dbCAN_2:Hotpep | 0.005626473 | 0.000733273 | 0.010519674 | 0.009240632 |
| CUPP | dbCAN_3 | -0.004837709 | -0.009730909 | 5.55E-05 | 0.056067044 |
| CUPP | dbCAN_4:dbCANsub | -0.004631537 | -0.009524737 | 0.000261664 | 0.084314612 |
| CUPP | dbCAN_4:DIAMOND | -0.004602789 | -0.00949599 | 0.000290411 | 0.089050932 |
| CUPP | dbCAN_4 | -0.004415464 | -0.009308664 | 0.000477736 | 0.125401127 |
| CUPP | dbCAN_3:HMMER | -0.004410112 | -0.009303312 | 0.000483089 | 0.126588368 |
| CUPP | dbCAN_4:HMMER | -0.004410112 | -0.009303312 | 0.000483089 | 0.126588368 |
| CUPP | dbCAN_3:DIAMOND | -0.004389312 | -0.009282512 | 0.000503889 | 0.131284483 |
| dbCAN_3:eCAMI | dbCAN_2:HMMER | -0.004231409 | -0.009099058 | 0.000636239 | 0.165244434 |
| dbCAN_3:eCAMI | dbCAN_2 | -0.003816016 | -0.008683664 | 0.001051632 | 0.307896084 |
| dbCAN_3 | dbCAN_2:DIAMOND | 0.003558143 | -0.001335058 | 0.008451343 | 0.432101325 |
| dbCAN_3:eCAMI | dbCAN_2:DIAMOND | -0.003392551 | -0.008260199 | 0.001475098 | 0.504445088 |
| dbCAN_2:Hotpep | dbCAN_2:HMMER | -0.003355586 | -0.008248787 | 0.001537614 | 0.531818285 |
| dbCAN_4:dbCANsub | dbCAN_2:DIAMOND | 0.003351971 | -0.00154123 | 0.008245171 | 0.533630716 |
| dbCAN_4:DIAMOND | dbCAN_2:DIAMOND | 0.003323223 | -0.001569977 | 0.008216423 | 0.548053012 |
| dbCAN_4 | dbCAN_2:DIAMOND | 0.003135898 | -0.001757302 | 0.008029098 | 0.641405084 |
| dbCAN_3 | dbCAN_2 | 0.003134677 | -0.001758523 | 0.008027878 | 0.642003029 |
| dbCAN_3:HMMER | dbCAN_2:DIAMOND | 0.003130545 | -0.001762655 | 0.008023746 | 0.644025281 |
| dbCAN_4:HMMER | dbCAN_2:DIAMOND | 0.003130545 | -0.001762655 | 0.008023746 | 0.644025281 |
| dbCAN_3:DIAMOND | dbCAN_2:DIAMOND | 0.003109746 | -0.001783455 | 0.008002946 | 0.654166497 |
| dbCAN_2:Hotpep | dbCAN_2 | -0.002940193 | -0.007833393 | 0.001953007 | 0.733574765 |
| dbCAN_4:dbCANsub | dbCAN_2 | 0.002928505 | -0.001964695 | 0.007821705 | 0.7387804 |
| dbCAN_4:DIAMOND | dbCAN_2 | 0.002899758 | -0.001993443 | 0.007792958 | 0.75140777 |
| dbCAN_3 | dbCAN_2:HMMER | 0.002719284 | -0.002173917 | 0.007612484 | 0.824117906 |
| dbCAN_4 | dbCAN_2 | 0.002712432 | -0.002180768 | 0.007605633 | 0.82663106 |
| dbCAN_3:HMMER | dbCAN_2 | 0.00270708 | -0.00218612 | 0.00760028 | 0.828580955 |
| dbCAN_4:HMMER | dbCAN_2 | 0.00270708 | -0.00218612 | 0.00760028 | 0.828580955 |
| dbCAN_3:DIAMOND | dbCAN_2 | 0.00268628 | -0.00220692 | 0.007579481 | 0.836044873 |
| dbCAN_2:Hotpep | dbCAN_2:DIAMOND | -0.002516728 | -0.007409928 | 0.002376473 | 0.88989792 |
| dbCAN_4:dbCANsub | dbCAN_2:HMMER | 0.002513112 | -0.002380089 | 0.007406312 | 0.890907538 |
| dbCAN_4:DIAMOND | dbCAN_2:HMMER | 0.002484364 | -0.002408836 | 0.007377565 | 0.898726079 |
| dbCAN_4 | dbCAN_2:HMMER | 0.002297039 | -0.002596161 | 0.007190239 | 0.940786384 |
| dbCAN_3:HMMER | dbCAN_2:HMMER | 0.002291687 | -0.002601514 | 0.007184887 | 0.941769558 |
| dbCAN_4:HMMER | dbCAN_2:HMMER | 0.002291687 | -0.002601514 | 0.007184887 | 0.941769558 |
| dbCAN_3:DIAMOND | dbCAN_2:HMMER | 0.002270887 | -0.002622314 | 0.007164087 | 0.945480558 |
| CUPP | dbCAN_3:eCAMI | 0.002112984 | -0.002754664 | 0.006980633 | 0.967003751 |
| CUPP | dbCAN_2:HMMER | -0.002118425 | -0.007011625 | 0.002774775 | 0.967679886 |
| CUPP | dbCAN_2 | -0.001703032 | -0.006596232 | 0.003190169 | 0.994933922 |
| CUPP | dbCAN_2:DIAMOND | -0.001279566 | -0.006172767 | 0.003613634 | 0.999689497 |
| CUPP | dbCAN_2:Hotpep | 0.001237161 | -0.003656039 | 0.006130362 | 0.999781029 |
| dbCAN_3:eCAMI | dbCAN_2:Hotpep | -0.000875823 | -0.005743471 | 0.003991825 | 0.999994511 |
| dbCAN_2:DIAMOND | dbCAN_2:HMMER | -0.000838859 | -0.005732059 | 0.004054342 | 0.999996809 |
| dbCAN_3:DIAMOND | dbCAN_3 | -0.000448397 | -0.005341597 | 0.004444803 | 0.999999998 |
| dbCAN_3:HMMER | dbCAN_3 | -0.000427597 | -0.005320798 | 0.004465603 | 0.999999999 |
| dbCAN_4:HMMER | dbCAN_3 | -0.000427597 | -0.005320798 | 0.004465603 | 0.999999999 |
| dbCAN_2:DIAMOND | dbCAN_2 | -0.000423465 | -0.005316666 | 0.004469735 | 0.999999999 |
| dbCAN_4 | dbCAN_3 | -0.000422245 | -0.005315445 | 0.004470956 | 0.999999999 |
| dbCAN_2:HMMER | dbCAN_2 | 0.000415393 | -0.004477807 | 0.005308594 | 0.999999999 |
| dbCAN_4:dbCANsub | dbCAN_3:DIAMOND | 0.000242225 | -0.004650975 | 0.005135425 | 1 |
| dbCAN_4:DIAMOND | dbCAN_3 | -0.00023492 | -0.00512812 | 0.004658281 | 1 |
| dbCAN_4:dbCANsub | dbCAN_3:HMMER | 0.000221425 | -0.004671775 | 0.005114625 | 1 |
| dbCAN_4:dbCANsub | dbCAN_4:HMMER | 0.000221425 | -0.004671775 | 0.005114625 | 1 |
| dbCAN_4:dbCANsub | dbCAN_3 | -0.000206172 | -0.005099372 | 0.004687028 | 1 |
| dbCAN_3:DIAMOND | dbCAN_3:HMMER | -2.08E-05 | -0.004914 | 0.004872401 | 1 |
| dbCAN_4 | dbCAN_3:HMMER | 5.35E-06 | -0.004887848 | 0.004898553 | 1 |
| dbCAN_4:HMMER | dbCAN_3:HMMER | 0 | -0.0048932 | 0.0048932 | 1 |
| dbCAN_4:DIAMOND | dbCAN_3:HMMER | 0.000192678 | -0.004700523 | 0.005085878 | 1 |
| dbCAN_4 | dbCAN_3:DIAMOND | 2.62E-05 | -0.004867048 | 0.004919353 | 1 |
| dbCAN_4:HMMER | dbCAN_3:DIAMOND | 2.08E-05 | -0.004872401 | 0.004914 | 1 |
| dbCAN_4:DIAMOND | dbCAN_3:DIAMOND | 0.000213477 | -0.004679723 | 0.005106678 | 1 |
| dbCAN_4:HMMER | dbCAN_4 | -5.35E-06 | -0.004898553 | 0.004887848 | 1 |
| dbCAN_4:DIAMOND | dbCAN_4 | 0.000187325 | -0.004705875 | 0.005080525 | 1 |
| dbCAN_4:dbCANsub | dbCAN_4 | 0.000216073 | -0.004677128 | 0.005109273 | 1 |
| dbCAN_4:DIAMOND | dbCAN_4:HMMER | 0.000192678 | -0.004700523 | 0.005085878 | 1 |
| dbCAN_4:dbCANsub | dbCAN_4:DIAMOND | 2.87E-05 | -0.004864453 | 0.004921948 | 1 |

**SI table 40: Output of Tukey HSD test for statistically significant differences between the mean sensitivity of tools classifying PL CAZyme domains (overleaf)**

| Classifier 1 | Classifier 2 | Mean Difference | Lower 95% Confidence Interval | Upper 95% Confidence Interval | Adjusted P-value |
|---|---|---|---|---|---|
| dbCAN_4:DIAMOND | dbCAN_3:eCAMI | 0.202708179 | 0.077541343 | 0.327875016 | 7.89E-06 |
| dbCAN_3:eCAMI | dbCAN_3:DIAMOND | -0.192069881 | -0.317236718 | -0.066903045 | 3.38E-05 |
| dbCAN_3:eCAMI | dbCAN_3 | -0.192069881 | -0.317236718 | -0.066903045 | 3.38E-05 |
| dbCAN_4:dbCANsub | dbCAN_3:eCAMI | 0.179137049 | 0.053970212 | 0.304303885 | 0.000179447 |
| dbCAN_3:eCAMI | dbCAN_3:HMMER | -0.177885484 | -0.303052321 | -0.052718648 | 0.00020965 |
| dbCAN_4:HMMER | dbCAN_3:eCAMI | 0.177885484 | 0.052718648 | 0.303052321 | 0.00020965 |
| dbCAN_4 | dbCAN_3:eCAMI | 0.177885484 | 0.052718648 | 0.303052321 | 0.00020965 |
| dbCAN_4:DIAMOND | dbCAN_2:Hotpep | 0.158010261 | 0.032186376 | 0.283834146 | 0.002338031 |
| CUPP | dbCAN_4:DIAMOND | -0.147673928 | -0.273497812 | -0.021850043 | 0.006870567 |
| dbCAN_3:DIAMOND | dbCAN_2:Hotpep | 0.147371963 | 0.021548078 | 0.273195848 | 0.007081338 |
| dbCAN_3 | dbCAN_2:Hotpep | 0.147371963 | 0.021548078 | 0.273195848 | 0.007081338 |
| CUPP | dbCAN_3 | -0.13703563 | -0.262859515 | -0.011211745 | 0.019029625 |
| CUPP | dbCAN_3:DIAMOND | -0.13703563 | -0.262859515 | -0.011211745 | 0.019029625 |
| dbCAN_4:dbCANsub | dbCAN_2:Hotpep | 0.13443913 | 0.008615246 | 0.260263015 | 0.024043894 |
| dbCAN_4:HMMER | dbCAN_2:Hotpep | 0.133187566 | 0.007363681 | 0.259011451 | 0.026855541 |
| dbCAN_4 | dbCAN_2:Hotpep | 0.133187566 | 0.007363681 | 0.259011451 | 0.026855541 |
| dbCAN_3:HMMER | dbCAN_2:Hotpep | 0.133187566 | 0.007363681 | 0.259011451 | 0.026855541 |
| dbCAN_4:DIAMOND | dbCAN_2:DIAMOND | 0.129662574 | 0.00383869 | 0.255486459 | 0.036390207 |
| CUPP | dbCAN_4:dbCANsub | -0.124102797 | -0.249926682 | 0.001721088 | 0.057392531 |
| CUPP | dbCAN_4:HMMER | -0.122851233 | -0.248675117 | 0.002972652 | 0.063331358 |
| CUPP | dbCAN_4 | -0.122851233 | -0.248675117 | 0.002972652 | 0.063331358 |
| CUPP | dbCAN_3:HMMER | -0.122851233 | -0.248675117 | 0.002972652 | 0.063331358 |
| dbCAN_3 | dbCAN_2:DIAMOND | 0.119024276 | -0.006799608 | 0.244848161 | 0.084761567 |
| dbCAN_3:DIAMOND | dbCAN_2:DIAMOND | 0.119024276 | -0.006799608 | 0.244848161 | 0.084761567 |
| dbCAN_4:DIAMOND | dbCAN_2 | 0.119024276 | -0.006799608 | 0.244848161 | 0.084761567 |
| dbCAN_3 | dbCAN_2 | 0.108385979 | -0.017437906 | 0.234209863 | 0.17598337 |
| dbCAN_3:DIAMOND | dbCAN_2 | 0.108385979 | -0.017437906 | 0.234209863 | 0.17598337 |
| dbCAN_4:dbCANsub | dbCAN_2:DIAMOND | 0.106091444 | -0.019732441 | 0.231915328 | 0.20272789 |
| dbCAN_4:HMMER | dbCAN_2:DIAMOND | 0.104839879 | -0.020984006 | 0.230663764 | 0.218444818 |
| dbCAN_4 | dbCAN_2:DIAMOND | 0.104839879 | -0.020984006 | 0.230663764 | 0.218444818 |
| dbCAN_3:HMMER | dbCAN_2:DIAMOND | 0.104839879 | -0.020984006 | 0.230663764 | 0.218444818 |
| dbCAN_3:eCAMI | dbCAN_2:HMMER | -0.101414399 | -0.226581236 | 0.023752437 | 0.25782352 |
| dbCAN_4:DIAMOND | dbCAN_2:HMMER | 0.10129378 | -0.024530105 | 0.227117665 | 0.267292379 |
| dbCAN_4:dbCANsub | dbCAN_2 | 0.095453146 | -0.030370739 | 0.221277031 | 0.360879222 |
| dbCAN_4:HMMER | dbCAN_2 | 0.094201581 | -0.031622303 | 0.220025466 | 0.382813752 |
| dbCAN_4 | dbCAN_2 | 0.094201581 | -0.031622303 | 0.220025466 | 0.382813752 |
| dbCAN_3:HMMER | dbCAN_2 | 0.094201581 | -0.031622303 | 0.220025466 | 0.382813752 |
| dbCAN_3 | dbCAN_2:HMMER | 0.090655482 | -0.035168403 | 0.216479367 | 0.44780383 |
| dbCAN_3:DIAMOND | dbCAN_2:HMMER | 0.090655482 | -0.035168403 | 0.216479367 | 0.44780383 |
| dbCAN_3:eCAMI | dbCAN_2 | -0.083683903 | -0.208850739 | 0.041482934 | 0.574018828 |
| dbCAN_4:dbCANsub | dbCAN_2:HMMER | 0.077722649 | -0.048101235 | 0.203546534 | 0.695816582 |
| dbCAN_4:HMMER | dbCAN_2:HMMER | 0.076471085 | -0.0493528 | 0.20229497 | 0.718340556 |
| dbCAN_4 | dbCAN_2:HMMER | 0.076471085 | -0.0493528 | 0.20229497 | 0.718340556 |
| dbCAN_3:HMMER | dbCAN_2:HMMER | 0.076471085 | -0.0493528 | 0.20229497 | 0.718340556 |
| dbCAN_3:eCAMI | dbCAN_2:DIAMOND | -0.073045605 | -0.198212441 | 0.052121232 | 0.770285222 |
| dbCAN_2:Hotpep | dbCAN_2:HMMER | -0.056716481 | -0.182540366 | 0.069107404 | 0.956019878 |
| CUPP | dbCAN_3:eCAMI | 0.055034251 | -0.070132585 | 0.180201088 | 0.963573846 |
| CUPP | dbCAN_2:HMMER | -0.046380148 | -0.172204033 | 0.079443737 | 0.991518736 |
| dbCAN_3:eCAMI | dbCAN_2:Hotpep | -0.044697918 | -0.169864755 | 0.080468918 | 0.993604844 |
| dbCAN_2:Hotpep | dbCAN_2 | -0.038985984 | -0.164809869 | 0.0868379 | 0.998303554 |
| CUPP | dbCAN_2 | -0.028649651 | -0.154473536 | 0.097174233 | 0.999927557 |
| dbCAN_2:DIAMOND | dbCAN_2:HMMER | -0.028368794 | -0.154192679 | 0.09745509 | 0.999934814 |
| dbCAN_2:Hotpep | dbCAN_2:DIAMOND | -0.028347687 | -0.154171571 | 0.097476198 | 0.999935332 |
| dbCAN_4:DIAMOND | dbCAN_3:HMMER | 0.024822695 | -0.10100119 | 0.15064658 | 0.999984747 |
| dbCAN_4:DIAMOND | dbCAN_4 | 0.024822695 | -0.10100119 | 0.15064658 | 0.999984747 |
| dbCAN_4:DIAMOND | dbCAN_4:HMMER | 0.024822695 | -0.10100119 | 0.15064658 | 0.999984747 |
| dbCAN_4:dbCANsub | dbCAN_4:DIAMOND | -0.023571131 | -0.149395015 | 0.102252754 | 0.999991395 |
| CUPP | dbCAN_2:DIAMOND | -0.018011353 | -0.143835238 | 0.107812531 | 0.999999589 |
| dbCAN_2:HMMER | dbCAN_2 | 0.017730496 | -0.108093388 | 0.143554381 | 0.999999657 |
| dbCAN_3:HMMER | dbCAN_3 | -0.014184397 | -0.140008282 | 0.111639488 | 0.999999974 |
| dbCAN_4 | dbCAN_3 | -0.014184397 | -0.140008282 | 0.111639488 | 0.999999974 |
| dbCAN_4:HMMER | dbCAN_3 | -0.014184397 | -0.140008282 | 0.111639488 | 0.999999974 |
| dbCAN_3:DIAMOND | dbCAN_3:HMMER | 0.014184397 | -0.111639488 | 0.140008282 | 0.999999974 |
| dbCAN_4 | dbCAN_3:DIAMOND | -0.014184397 | -0.140008282 | 0.111639488 | 0.999999974 |
| dbCAN_4:HMMER | dbCAN_3:DIAMOND | -0.014184397 | -0.140008282 | 0.111639488 | 0.999999974 |
| dbCAN_4:dbCANsub | dbCAN_3 | -0.012932833 | -0.138756717 | 0.112891052 | 0.999999991 |
| dbCAN_4:dbCANsub | dbCAN_3:DIAMOND | -0.012932833 | -0.138756717 | 0.112891052 | 0.999999991 |
| dbCAN_2:DIAMOND | dbCAN_2 | -0.010638298 | -0.136462183 | 0.115185587 | 0.999999999 |
| dbCAN_4:DIAMOND | dbCAN_3 | 0.010638298 | -0.115185587 | 0.136462183 | 0.999999999 |
| dbCAN_4:DIAMOND | dbCAN_3:DIAMOND | 0.010638298 | -0.115185587 | 0.136462183 | 0.999999999 |
| CUPP | dbCAN_2:Hotpep | 0.010336333 | -0.115487552 | 0.136160218 | 0.999999999 |
| dbCAN_3:DIAMOND | dbCAN_3 | -2.22E-16 | -0.125823885 | 0.125823885 | 1 |
| dbCAN_4 | dbCAN_3:HMMER | 2.22E-16 | -0.125823885 | 0.125823885 | 1 |
| dbCAN_4:HMMER | dbCAN_3:HMMER | 4.44E-16 | -0.125823885 | 0.125823885 | 1 |
| dbCAN_4:dbCANsub | dbCAN_3:HMMER | 0.001251564 | -0.12457232 | 0.127075449 | 1 |
| dbCAN_4:HMMER | dbCAN_4 | 2.22E-16 | -0.125823885 | 0.125823885 | 1 |
| dbCAN_4:dbCANsub | dbCAN_4 | 0.001251564 | -0.12457232 | 0.127075449 | 1 |
| dbCAN_4:dbCANsub | dbCAN_4:HMMER | 0.001251564 | -0.12457232 | 0.127075449 | 1 |

### 6.4 Carbohydrate esterases

**SI table 41: Output of Tukey HSD test for statistically significant differences between the mean F1-score of tools classifying CE CAZyme domains (overleaf)**

| Classifier 1 | Classifier 2 | Mean Difference | Lower 95% Confidence Interval | Upper 95% Confidence Interval | Adjusted P-value |
|---|---|---|---|---|---|
| dbCAN_4 | dbCAN_3:eCAMI | 0.132425143 | 0.056418829 | 0.208431458 | 7.50E-07 |
| dbCAN_4:dbCANsub | dbCAN_3:eCAMI | 0.130346287 | 0.054339972 | 0.206352601 | 1.26E-06 |
| dbCAN_4:DIAMOND | dbCAN_3:eCAMI | 0.119310227 | 0.043303913 | 0.195316542 | 1.73E-05 |
| dbCAN_4:HMMER | dbCAN_3:eCAMI | 0.117346435 | 0.041340121 | 0.19335275 | 2.69E-05 |
| dbCAN_4 | dbCAN_2:Hotpep | 0.10833257 | 0.032326256 | 0.184338885 | 0.000187738 |
| dbCAN_4:dbCANsub | dbCAN_2:Hotpep | 0.106253714 | 0.030247399 | 0.182260028 | 0.000287475 |
| dbCAN_4 | dbCAN_2:DIAMOND | 0.100138788 | 0.024132474 | 0.176145103 | 0.000958046 |
| dbCAN_3:eCAMI | dbCAN_3:HMMER | -0.098854622 | -0.174860937 | -0.022848308 | 0.001221778 |
| dbCAN_4:dbCANsub | dbCAN_2:DIAMOND | 0.098059932 | 0.022053617 | 0.174066246 | 0.001417758 |
| dbCAN_4:DIAMOND | dbCAN_2:Hotpep | 0.095217654 | 0.01921134 | 0.171223969 | 0.002387627 |
| CUPP | dbCAN_3:eCAMI | 0.093651409 | 0.017645094 | 0.169657723 | 0.003158667 |
| dbCAN_3:eCAMI | dbCAN_2:HMMER | -0.093481815 | -0.16948813 | -0.017475501 | 0.003254809 |
| dbCAN_4:HMMER | dbCAN_2:Hotpep | 0.093253862 | 0.017247548 | 0.169260177 | 0.003388326 |
| dbCAN_3:eCAMI | dbCAN_3 | -0.092221131 | -0.168227446 | -0.016214817 | 0.004059445 |
| dbCAN_3:eCAMI | dbCAN_2 | -0.091038008 | -0.167044322 | -0.015031693 | 0.004978752 |
| dbCAN_4:DIAMOND | dbCAN_2:DIAMOND | 0.087023872 | 0.011017558 | 0.163030187 | 0.009720562 |
| dbCAN_4:HMMER | dbCAN_2:DIAMOND | 0.08506008 | 0.009053766 | 0.161066395 | 0.013302878 |
| dbCAN_3:eCAMI | dbCAN_3:DIAMOND | -0.083883002 | -0.159889316 | -0.007876687 | 0.015984799 |
| dbCAN_3:HMMER | dbCAN_2:Hotpep | 0.074762049 | -0.001244265 | 0.150768364 | 0.058994294 |
| CUPP | dbCAN_2:Hotpep | 0.069558836 | -0.006447479 | 0.14556515 | 0.112484075 |
| dbCAN_2:Hotpep | dbCAN_2:HMMER | -0.069389242 | -0.145395557 | 0.006617072 | 0.114726309 |
| dbCAN_3 | dbCAN_2:Hotpep | 0.068128559 | -0.007877756 | 0.144134873 | 0.132505685 |
| dbCAN_2:Hotpep | dbCAN_2 | -0.066945435 | -0.14295175 | 0.009060879 | 0.151040238 |
| dbCAN_3:HMMER | dbCAN_2:DIAMOND | 0.066568267 | -0.009438047 | 0.142574582 | 0.157338264 |
| CUPP | dbCAN_2:DIAMOND | 0.061365054 | -0.014641261 | 0.137371368 | 0.264291338 |
| dbCAN_2:DIAMOND | dbCAN_2:HMMER | -0.06119546 | -0.137201775 | 0.014810854 | 0.268409673 |
| dbCAN_3 | dbCAN_2:DIAMOND | 0.059934776 | -0.016071538 | 0.135941091 | 0.300223986 |
| dbCAN_3:DIAMOND | dbCAN_2:Hotpep | 0.059790429 | -0.016215885 | 0.135796744 | 0.303998719 |
| dbCAN_2:DIAMOND | dbCAN_2 | -0.058751653 | -0.134757968 | 0.017254661 | 0.331928188 |
| dbCAN_3:DIAMOND | dbCAN_2:DIAMOND | 0.051596647 | -0.024409667 | 0.127602962 | 0.551142796 |
| dbCAN_4 | dbCAN_3:DIAMOND | 0.048542141 | -0.027464173 | 0.124548456 | 0.64908784 |
| dbCAN_4:dbCANsub | dbCAN_3:DIAMOND | 0.046463285 | -0.02954303 | 0.122469599 | 0.712697453 |
| dbCAN_4 | dbCAN_2 | 0.041387135 | -0.034619179 | 0.11739345 | 0.845463819 |
| dbCAN_4 | dbCAN_3 | 0.040204012 | -0.035802303 | 0.116210326 | 0.870327825 |
| dbCAN_4:dbCANsub | dbCAN_2 | 0.039308279 | -0.036698036 | 0.115314593 | 0.887442635 |
| dbCAN_4 | dbCAN_2:HMMER | 0.038943328 | -0.037062987 | 0.114949642 | 0.893989014 |
| CUPP | dbCAN_4 | -0.038773734 | -0.114780049 | 0.03723258 | 0.896946987 |
| dbCAN_4:dbCANsub | dbCAN_3 | 0.038125155 | -0.037881159 | 0.11413147 | 0.907768049 |
| dbCAN_4:dbCANsub | dbCAN_2:HMMER | 0.036864472 | -0.039141843 | 0.112870786 | 0.926600055 |
| CUPP | dbCAN_4:dbCANsub | -0.036694878 | -0.112701192 | 0.039311437 | 0.928915554 |
| dbCAN_4:DIAMOND | dbCAN_3:DIAMOND | 0.035427225 | -0.040579089 | 0.11143354 | 0.944643607 |
| dbCAN_4 | dbCAN_3:HMMER | 0.033570521 | -0.042435794 | 0.109576835 | 0.962954984 |
| dbCAN_4:HMMER | dbCAN_3:DIAMOND | 0.033463433 | -0.042542881 | 0.109469748 | 0.963852443 |
| dbCAN_3:eCAMI | dbCAN_2:DIAMOND | -0.032286355 | -0.108292669 | 0.04371996 | 0.972677655 |
| dbCAN_4:dbCANsub | dbCAN_3:HMMER | 0.031491665 | -0.04451465 | 0.107497979 | 0.977632934 |
| dbCAN_4:DIAMOND | dbCAN_2 | 0.028272219 | -0.047734095 | 0.104278534 | 0.991017886 |
| dbCAN_4:DIAMOND | dbCAN_3 | 0.027089096 | -0.048917219 | 0.10309541 | 0.993862272 |
| dbCAN_4:HMMER | dbCAN_2 | 0.026308427 | -0.049697887 | 0.102314742 | 0.995296603 |
| dbCAN_4:DIAMOND | dbCAN_2:HMMER | 0.025828412 | -0.050177903 | 0.101834726 | 0.996031415 |
| CUPP | dbCAN_4:DIAMOND | -0.025658818 | -0.101665133 | 0.050347496 | 0.99626696 |
| dbCAN_4:HMMER | dbCAN_3 | 0.025125304 | -0.050881011 | 0.101131618 | 0.99693315 |
| dbCAN_3:eCAMI | dbCAN_2:Hotpep | -0.024092573 | -0.100098887 | 0.051913742 | 0.997941917 |
| dbCAN_4:HMMER | dbCAN_2:HMMER | 0.02386462 | -0.052141694 | 0.099870935 | 0.998121781 |
| CUPP | dbCAN_4:HMMER | -0.023695026 | -0.099701341 | 0.052311288 | 0.998246773 |
| dbCAN_4:DIAMOND | dbCAN_3:HMMER | 0.020455605 | -0.05555071 | 0.096461919 | 0.999594761 |
| dbCAN_4:HMMER | dbCAN_3:HMMER | 0.018491813 | -0.057514501 | 0.094498128 | 0.999858014 |
| dbCAN_4:HMMER | dbCAN_4 | -0.015078708 | -0.091085022 | 0.060927607 | 0.999984302 |
| dbCAN_3:DIAMOND | dbCAN_3:HMMER | -0.01497162 | -0.090977935 | 0.061034694 | 0.999985488 |
| dbCAN_4:DIAMOND | dbCAN_4 | -0.013114916 | -0.089121231 | 0.062891398 | 0.999996682 |
| dbCAN_4:dbCANsub | dbCAN_4:HMMER | 0.012999851 | -0.063006463 | 0.089006166 | 0.999996995 |
| dbCAN_4:dbCANsub | dbCAN_4:DIAMOND | 0.01103606 | -0.064970255 | 0.087042374 | 0.999999533 |
| CUPP | dbCAN_3:DIAMOND | 0.009768407 | -0.066237908 | 0.085774721 | 0.999999886 |
| dbCAN_3:DIAMOND | dbCAN_2:HMMER | -0.009598813 | -0.085605128 | 0.066407501 | 0.999999907 |
| dbCAN_3:DIAMOND | dbCAN_3 | -0.008338129 | -0.084344444 | 0.067668185 | 0.999999982 |
| dbCAN_2:Hotpep | dbCAN_2:DIAMOND | -0.008193782 | -0.084200097 | 0.067812532 | 0.999999985 |
| dbCAN_3:HMMER | dbCAN_2 | 0.007816614 | -0.0681897 | 0.083822929 | 0.999999991 |
| dbCAN_3:DIAMOND | dbCAN_2 | -0.007155006 | -0.083161321 | 0.068851308 | 0.999999997 |
| dbCAN_3:HMMER | dbCAN_3 | 0.006633491 | -0.069372824 | 0.082639805 | 0.999999999 |
| dbCAN_3:HMMER | dbCAN_2:HMMER | 0.005372807 | -0.070633507 | 0.081379122 | 1 |
| CUPP | dbCAN_3:HMMER | -0.005203213 | -0.081209528 | 0.070803101 | 1 |
| dbCAN_2:HMMER | dbCAN_2 | 0.002443807 | -0.073562507 | 0.078450122 | 1 |
| dbCAN_3 | dbCAN_2 | 0.001183123 | -0.074823191 | 0.077189438 | 1 |
| CUPP | dbCAN_2 | 0.002613401 | -0.073392914 | 0.078619715 | 1 |
| dbCAN_3 | dbCAN_2:HMMER | -0.001260684 | -0.077266998 | 0.074745631 | 1 |
| CUPP | dbCAN_2:HMMER | 0.000169594 | -0.075836721 | 0.076175908 | 1 |
| CUPP | dbCAN_3 | 0.001430277 | -0.074576037 | 0.077436592 | 1 |
| dbCAN_4:dbCANsub | dbCAN_4 | -0.002078856 | -0.078085171 | 0.073927458 | 1 |
| dbCAN_4:DIAMOND | dbCAN_4:HMMER | 0.001963792 | -0.074042523 | 0.077970106 | 1 |

**SI table 42: Output of Tukey HSD test for statistically significant differences between the mean accuracy of tools classifying CE CAZyme domains (overleaf)**

| Classifier 1 | Classifier 2 | Mean Difference | Lower 95% Confidence Interval | Upper 95% Confidence Interval | Adjusted P-value |
|---|---|---|---|---|---|
| dbCAN_4 | dbCAN_3:eCAMI | 0.012516122 | 0.001187881 | 0.023844363 | 0.015755589 |
| dbCAN_4 | dbCAN_2:Hotpep | 0.012199394 | 0.000871153 | 0.023527635 | 0.021774505 |
| dbCAN_4:dbCANsub | dbCAN_3:eCAMI | 0.012019837 | 0.000691596 | 0.023348078 | 0.026028993 |
| dbCAN_4:dbCANsub | dbCAN_2:Hotpep | 0.011703109 | 0.000374868 | 0.02303135 | 0.035342759 |
| dbCAN_4:HMMER | dbCAN_3:eCAMI | 0.01106777 | -0.000260471 | 0.022396011 | 0.063002129 |
| dbCAN_4:HMMER | dbCAN_2:Hotpep | 0.010751042 | -0.000577199 | 0.022079283 | 0.082514192 |
| dbCAN_4:DIAMOND | dbCAN_3:eCAMI | 0.010148267 | -0.001179974 | 0.021476508 | 0.133090976 |
| dbCAN_4:DIAMOND | dbCAN_2:Hotpep | 0.009831539 | -0.001496702 | 0.02115978 | 0.167824681 |
| dbCAN_3:eCAMI | dbCAN_3:HMMER | -0.009235935 | -0.020564176 | 0.002092306 | 0.249965083 |
| dbCAN_3:HMMER | dbCAN_2:Hotpep | 0.008919207 | -0.002409034 | 0.020247448 | 0.30262063 |
| dbCAN_3:eCAMI | dbCAN_2:HMMER | -0.008349105 | -0.019677346 | 0.002979136 | 0.411201146 |
| CUPP | dbCAN_3:eCAMI | 0.008211867 | -0.003116374 | 0.019540108 | 0.439466517 |
| dbCAN_4 | dbCAN_2:DIAMOND | 0.008075706 | -0.003252535 | 0.019403947 | 0.46809803 |
| dbCAN_2:Hotpep | dbCAN_2:HMMER | -0.008032377 | -0.019360618 | 0.003295864 | 0.477310658 |
| CUPP | dbCAN_2:Hotpep | 0.007895139 | -0.003433102 | 0.01922338 | 0.506740963 |
| dbCAN_3:eCAMI | dbCAN_3 | -0.007799184 | -0.019127425 | 0.003529057 | 0.527483683 |
| dbCAN_4:dbCANsub | dbCAN_2:DIAMOND | 0.007579421 | -0.00374882 | 0.018907662 | 0.575178859 |
| dbCAN_3:eCAMI | dbCAN_2 | -0.007523169 | -0.01885141 | 0.003805072 | 0.587367584 |
| dbCAN_3 | dbCAN_2:Hotpep | 0.007482456 | -0.003845785 | 0.018810697 | 0.596170677 |
| dbCAN_2:Hotpep | dbCAN_2 | -0.007206441 | -0.018534682 | 0.0041218 | 0.655073015 |
| dbCAN_3:eCAMI | dbCAN_3:DIAMOND | -0.007204837 | -0.018533078 | 0.004123404 | 0.65540984 |
| dbCAN_3:DIAMOND | dbCAN_2:Hotpep | 0.006888109 | -0.004440132 | 0.01821635 | 0.720005532 |
| dbCAN_4:HMMER | dbCAN_2:DIAMOND | 0.006627354 | -0.004700887 | 0.017955595 | 0.769466751 |
| dbCAN_4:DIAMOND | dbCAN_2:DIAMOND | 0.005707851 | -0.00562039 | 0.017036092 | 0.904983689 |
| dbCAN_4 | dbCAN_3:DIAMOND | 0.005311285 | -0.006016956 | 0.016639526 | 0.942243152 |
| dbCAN_4 | dbCAN_2 | 0.004992953 | -0.006335288 | 0.016321194 | 0.96354793 |
| dbCAN_4:dbCANsub | dbCAN_3:DIAMOND | 0.004815 | -0.006513241 | 0.016143241 | 0.972545178 |
| dbCAN_3:HMMER | dbCAN_2:DIAMOND | 0.004795519 | -0.006532722 | 0.01612376 | 0.973416353 |
| dbCAN_4 | dbCAN_3 | 0.004716938 | -0.006611303 | 0.016045179 | 0.976717258 |
| dbCAN_4:dbCANsub | dbCAN_2 | 0.004496668 | -0.006831573 | 0.015824909 | 0.984300122 |
| dbCAN_3:eCAMI | dbCAN_2:DIAMOND | -0.004440416 | -0.015768657 | 0.006887825 | 0.985881698 |
| CUPP | dbCAN_4 | -0.004304255 | -0.015632496 | 0.007023986 | 0.989188755 |
| dbCAN_4:dbCANsub | dbCAN_3 | 0.004220654 | -0.007107587 | 0.015548895 | 0.990888539 |
| dbCAN_4 | dbCAN_2:HMMER | 0.004167017 | -0.007161224 | 0.015495258 | 0.991859938 |
| dbCAN_2:Hotpep | dbCAN_2:DIAMOND | -0.004123688 | -0.015451929 | 0.007204553 | 0.992581503 |
| dbCAN_2:DIAMOND | dbCAN_2:HMMER | -0.003908689 | -0.01523693 | 0.007419552 | 0.995431581 |
| dbCAN_4:HMMER | dbCAN_3:DIAMOND | 0.003862933 | -0.007465308 | 0.015191174 | 0.995901813 |
| CUPP | dbCAN_4:dbCANsub | -0.003807971 | -0.015136211 | 0.00752027 | 0.996412633 |
| CUPP | dbCAN_2:DIAMOND | 0.003771451 | -0.00755679 | 0.015099692 | 0.996721683 |
| dbCAN_4:dbCANsub | dbCAN_2:HMMER | 0.003670732 | -0.007657509 | 0.014998973 | 0.997460648 |
| dbCAN_4:HMMER | dbCAN_2 | 0.003544601 | -0.00778364 | 0.014872842 | 0.998183339 |
| dbCAN_3 | dbCAN_2:DIAMOND | 0.003358768 | -0.007969473 | 0.014687009 | 0.998926499 |
| dbCAN_4 | dbCAN_3:HMMER | 0.003280187 | -0.008048054 | 0.014608428 | 0.999151294 |
| dbCAN_4:HMMER | dbCAN_3 | 0.003268586 | -0.008059655 | 0.014596827 | 0.999180772 |
| dbCAN_2:DIAMOND | dbCAN_2 | -0.003082753 | -0.014410994 | 0.008245488 | 0.99954619 |
| dbCAN_4:DIAMOND | dbCAN_3:DIAMOND | 0.00294343 | -0.008384811 | 0.014271671 | 0.999717788 |
| CUPP | dbCAN_4:HMMER | -0.002855903 | -0.014184144 | 0.008472338 | 0.999793781 |
| dbCAN_4:dbCANsub | dbCAN_3:HMMER | 0.002783902 | -0.008544339 | 0.014112143 | 0.999842181 |
| dbCAN_3:DIAMOND | dbCAN_2:DIAMOND | 0.002764421 | -0.00856382 | 0.014092662 | 0.999853422 |
| dbCAN_4:HMMER | dbCAN_2:HMMER | 0.002718665 | -0.008609576 | 0.014046905 | 0.999877099 |
| dbCAN_4:DIAMOND | dbCAN_2 | 0.002625098 | -0.008703143 | 0.013953339 | 0.999915287 |
| dbCAN_4:DIAMOND | dbCAN_4 | -0.002367855 | -0.013696096 | 0.008960386 | 0.999972173 |
| dbCAN_4:DIAMOND | dbCAN_3 | 0.002349084 | -0.008979157 | 0.013677324 | 0.999974489 |
| dbCAN_3:DIAMOND | dbCAN_3:HMMER | -0.002031098 | -0.013359339 | 0.009297143 | 0.999994898 |
| CUPP | dbCAN_4:DIAMOND | -0.0019364 | -0.013264641 | 0.00939184 | 0.999997015 |
| dbCAN_4:dbCANsub | dbCAN_4:DIAMOND | 0.00187157 | -0.009456671 | 0.013199811 | 0.999997968 |
| dbCAN_4:HMMER | dbCAN_3:HMMER | 0.001831835 | -0.009496406 | 0.013160076 | 0.999998406 |
| dbCAN_4:DIAMOND | dbCAN_2:HMMER | 0.001799162 | -0.009529079 | 0.013127403 | 0.999998701 |
| dbCAN_3:HMMER | dbCAN_2 | 0.001712766 | -0.009615475 | 0.013041007 | 0.999999258 |
| dbCAN_4:HMMER | dbCAN_4 | -0.001448352 | -0.012776593 | 0.009879889 | 0.999999892 |
| dbCAN_3:HMMER | dbCAN_3 | 0.001436751 | -0.00989149 | 0.012764992 | 0.999999902 |
| dbCAN_3:DIAMOND | dbCAN_2:HMMER | -0.001144268 | -0.012472509 | 0.010183973 | 0.999999993 |
| CUPP | dbCAN_3:HMMER | -0.001024068 | -0.012352309 | 0.010304173 | 0.999999998 |
| CUPP | dbCAN_3:DIAMOND | 0.00100703 | -0.010321211 | 0.012335271 | 0.999999998 |
| dbCAN_4:dbCANsub | dbCAN_4:HMMER | 0.000952067 | -0.010376173 | 0.012280308 | 0.999999999 |
| dbCAN_4:DIAMOND | dbCAN_4:HMMER | -0.000919503 | -0.012247744 | 0.010408738 | 0.999999999 |
| dbCAN_4:DIAMOND | dbCAN_3:HMMER | 0.000912332 | -0.010415909 | 0.012240573 | 1 |
| dbCAN_3:HMMER | dbCAN_2:HMMER | 0.00088683 | -0.010441411 | 0.012215071 | 1 |
| dbCAN_2:HMMER | dbCAN_2 | 0.000825936 | -0.010502305 | 0.012154177 | 1 |
| CUPP | dbCAN_2 | 0.000688698 | -0.010639543 | 0.012016938 | 1 |
| dbCAN_3:DIAMOND | dbCAN_3 | -0.000594347 | -0.011922588 | 0.010733894 | 1 |
| dbCAN_3 | dbCAN_2:HMMER | -0.000549922 | -0.011878163 | 0.010778319 | 1 |
| dbCAN_3 | dbCAN_2 | 0.000276014 | -0.011052226 | 0.011604255 | 1 |
| dbCAN_3:DIAMOND | dbCAN_2 | -0.000318332 | -0.011646573 | 0.011009909 | 1 |
| CUPP | dbCAN_2:HMMER | -0.000137239 | -0.011465479 | 0.011191002 | 1 |
| dbCAN_3:eCAMI | dbCAN_2:Hotpep | -0.000316728 | -0.011644969 | 0.011011513 | 1 |
| CUPP | dbCAN_3 | 0.000412683 | -0.010915558 | 0.011740924 | 1 |
| dbCAN_4:dbCANsub | dbCAN_4 | -0.000496285 | -0.011824526 | 0.010831956 | 1 |

**SI table 43: Output of Tukey HSD test for statistically significant differences between the mean sensitivity of tools classifying CE CAZyme domains (overleaf)**

| Classifier 1 | Classifier 2 | Mean Difference | Lower 95% Confidence Interval | Upper 95% Confidence Interval | Adjusted P-value |
|---|---|---|---|---|---|
| dbCAN_4 | dbCAN_3:eCAMI | 0.170937761 | 0.086770942 | 0.25510458 | 2.04E-09 |
| dbCAN_4:DIAMOND | dbCAN_3:eCAMI | 0.168170426 | 0.084003607 | 0.252337245 | 4.17E-09 |
| dbCAN_4:dbCANsub | dbCAN_3:eCAMI | 0.168144319 | 0.0839775 | 0.252311138 | 4.20E-09 |
| dbCAN_4:HMMER | dbCAN_3:eCAMI | 0.145577485 | 0.061410667 | 0.229744304 | 9.54E-07 |
| dbCAN_4 | dbCAN_2:DIAMOND | 0.130284043 | 0.046117225 | 0.214450862 | 2.51E-05 |
| dbCAN_4:DIAMOND | dbCAN_2:DIAMOND | 0.127516708 | 0.04334989 | 0.211683527 | 4.38E-05 |
| dbCAN_4:dbCANsub | dbCAN_2:DIAMOND | 0.127490602 | 0.043323783 | 0.21165742 | 4.40E-05 |
| dbCAN_4 | dbCAN_2:Hotpep | 0.127443609 | 0.04327679 | 0.211610428 | 4.44E-05 |
| dbCAN_4:DIAMOND | dbCAN_2:Hotpep | 0.124676274 | 0.040509455 | 0.208843093 | 7.65E-05 |
| dbCAN_4:dbCANsub | dbCAN_2:Hotpep | 0.124650167 | 0.040483348 | 0.208816986 | 7.69E-05 |
| dbCAN_3:eCAMI | dbCAN_3:DIAMOND | -0.122947995 | -0.207114814 | -0.038781176 | 0.00010673 |
| dbCAN_3:eCAMI | dbCAN_3 | -0.120958647 | -0.205125465 | -0.036791828 | 0.000155762 |
| dbCAN_3:eCAMI | dbCAN_3:HMMER | -0.11566416 | -0.199830979 | -0.031497342 | 0.000413235 |
| dbCAN_3:eCAMI | dbCAN_2 | -0.113977652 | -0.198144471 | -0.029810834 | 0.000558558 |
| dbCAN_3:eCAMI | dbCAN_2:HMMER | -0.113471178 | -0.197637997 | -0.029304359 | 0.000610909 |
| dbCAN_4:HMMER | dbCAN_2:DIAMOND | 0.104923768 | 0.020756949 | 0.189090587 | 0.002596797 |
| CUPP | dbCAN_3:eCAMI | 0.104114453 | 0.019947634 | 0.188281272 | 0.002958809 |
| dbCAN_4:HMMER | dbCAN_2:Hotpep | 0.102083333 | 0.017916514 | 0.186250152 | 0.004084515 |
| dbCAN_3:DIAMOND | dbCAN_2:DIAMOND | 0.082294277 | -0.001872541 | 0.166461096 | 0.062540102 |
| dbCAN_3 | dbCAN_2:DIAMOND | 0.080304929 | -0.00386189 | 0.164471748 | 0.078648798 |
| dbCAN_3:DIAMOND | dbCAN_2:Hotpep | 0.079453843 | -0.004712976 | 0.163620662 | 0.086513174 |
| dbCAN_3 | dbCAN_2:Hotpep | 0.077464495 | -0.006702324 | 0.161631313 | 0.10739543 |
| dbCAN_3:HMMER | dbCAN_2:DIAMOND | 0.075010443 | -0.009156376 | 0.159177262 | 0.138426426 |
| dbCAN_2:DIAMOND | dbCAN_2 | -0.073323935 | -0.157490754 | 0.010842884 | 0.163416192 |
| dbCAN_2:DIAMOND | dbCAN_2:HMMER | -0.07281746 | -0.156984279 | 0.011349359 | 0.171531559 |
| dbCAN_3:HMMER | dbCAN_2:Hotpep | 0.072170008 | -0.011996811 | 0.156336827 | 0.18232511 |
| dbCAN_2:Hotpep | dbCAN_2 | -0.0704835 | -0.154650319 | 0.013683318 | 0.212677504 |
| dbCAN_2:Hotpep | dbCAN_2:HMMER | -0.069977026 | -0.154143845 | 0.014189793 | 0.222427747 |
| CUPP | dbCAN_4 | -0.066823308 | -0.150990127 | 0.017343511 | 0.289658772 |
| CUPP | dbCAN_4:DIAMOND | -0.064055973 | -0.148222792 | 0.020110846 | 0.357340269 |
| CUPP | dbCAN_4:dbCANsub | -0.064029866 | -0.148196685 | 0.020136953 | 0.358013591 |
| CUPP | dbCAN_2:DIAMOND | 0.063460735 | -0.020706084 | 0.147627554 | 0.372839412 |
| CUPP | dbCAN_2:Hotpep | 0.060620301 | -0.023546518 | 0.14478712 | 0.450510012 |
| dbCAN_4 | dbCAN_2:HMMER | 0.057466583 | -0.026700236 | 0.141633402 | 0.541493632 |
| dbCAN_4 | dbCAN_2 | 0.056960109 | -0.02720671 | 0.141126927 | 0.55629505 |
| dbCAN_4 | dbCAN_3:HMMER | 0.055273601 | -0.028893218 | 0.13944042 | 0.605452229 |
| dbCAN_4:DIAMOND | dbCAN_2:HMMER | 0.054699248 | -0.029467571 | 0.138866067 | 0.622050744 |
| dbCAN_4:dbCANsub | dbCAN_2:HMMER | 0.054673141 | -0.029493678 | 0.13883996 | 0.622802532 |
| dbCAN_4:DIAMOND | dbCAN_2 | 0.054192774 | -0.029974045 | 0.138359592 | 0.636586532 |
| dbCAN_4:dbCANsub | dbCAN_2 | 0.054166667 | -0.030000152 | 0.138333486 | 0.637332815 |
| dbCAN_4:DIAMOND | dbCAN_3:HMMER | 0.052506266 | -0.031660553 | 0.136673085 | 0.684030772 |
| dbCAN_4:dbCANsub | dbCAN_3:HMMER | 0.052480159 | -0.03168666 | 0.136646978 | 0.684750987 |
| dbCAN_4 | dbCAN_3 | 0.049979114 | -0.034187704 | 0.134145933 | 0.751067267 |
| dbCAN_4 | dbCAN_3:DIAMOND | 0.047989766 | -0.036177053 | 0.132156585 | 0.799119584 |
| dbCAN_4:DIAMOND | dbCAN_3 | 0.047211779 | -0.036955039 | 0.131378598 | 0.816559604 |
| dbCAN_4:dbCANsub | dbCAN_3 | 0.047185673 | -0.036981146 | 0.131352491 | 0.817130729 |
| dbCAN_4:DIAMOND | dbCAN_3:DIAMOND | 0.045222431 | -0.038944388 | 0.12938925 | 0.857323988 |
| dbCAN_4:dbCANsub | dbCAN_3:DIAMOND | 0.045196324 | -0.038970495 | 0.129363143 | 0.857820818 |
| dbCAN_3:eCAMI | dbCAN_2:Hotpep | -0.043494152 | -0.127660971 | 0.040672667 | 0.888011219 |
| CUPP | dbCAN_4:HMMER | -0.041463033 | -0.125629851 | 0.042703786 | 0.918316347 |
| dbCAN_3:eCAMI | dbCAN_2:DIAMOND | -0.040653718 | -0.124820536 | 0.043513101 | 0.928682966 |
| dbCAN_4:HMMER | dbCAN_2:HMMER | 0.032106307 | -0.052060511 | 0.116273126 | 0.988813257 |
| dbCAN_4:HMMER | dbCAN_2 | 0.031599833 | -0.052566986 | 0.115766652 | 0.990256487 |
| dbCAN_4:HMMER | dbCAN_3:HMMER | 0.029913325 | -0.054253494 | 0.114080144 | 0.994016181 |
| dbCAN_4:HMMER | dbCAN_4 | -0.025360276 | -0.109527095 | 0.058806543 | 0.998741983 |
| dbCAN_4:HMMER | dbCAN_3 | 0.024618839 | -0.05954798 | 0.108785658 | 0.999061467 |
| dbCAN_4:HMMER | dbCAN_3:DIAMOND | 0.02262949 | -0.061537328 | 0.106796309 | 0.999598835 |
| dbCAN_4:DIAMOND | dbCAN_4:HMMER | 0.022592941 | -0.061573878 | 0.10675976 | 0.999605419 |
| dbCAN_4:dbCANsub | dbCAN_4:HMMER | 0.022566834 | -0.061599985 | 0.106733653 | 0.999610064 |
| CUPP | dbCAN_3:DIAMOND | -0.018833542 | -0.103000361 | 0.065333277 | 0.999941766 |
| CUPP | dbCAN_3 | -0.016844194 | -0.101011013 | 0.067322625 | 0.999982718 |
| CUPP | dbCAN_3:HMMER | -0.011549708 | -0.095716526 | 0.072617111 | 0.999999756 |
| CUPP | dbCAN_2 | -0.0098632 | -0.094030019 | 0.074303619 | 0.999999961 |
| dbCAN_3:DIAMOND | dbCAN_2:HMMER | 0.009476817 | -0.074690002 | 0.093643636 | 0.999999975 |
| CUPP | dbCAN_2:HMMER | -0.009356725 | -0.093523544 | 0.074810094 | 0.999999979 |
| dbCAN_3:DIAMOND | dbCAN_2 | 0.008970343 | -0.075196476 | 0.093137161 | 0.999999987 |
| dbCAN_3 | dbCAN_2:HMMER | 0.007487469 | -0.07667935 | 0.091654288 | 0.999999998 |
| dbCAN_3:DIAMOND | dbCAN_3:HMMER | 0.007283835 | -0.076882984 | 0.091450653 | 0.999999999 |
| dbCAN_3 | dbCAN_2 | 0.006980994 | -0.077185825 | 0.091147813 | 0.999999999 |
| dbCAN_3:HMMER | dbCAN_3 | -0.005294486 | -0.089461305 | 0.078872333 | 1 |
| dbCAN_2:HMMER | dbCAN_2 | -0.000506475 | -0.084673293 | 0.083660344 | 1 |
| dbCAN_3:HMMER | dbCAN_2 | 0.001686508 | -0.082480311 | 0.085853327 | 1 |
| dbCAN_3:HMMER | dbCAN_2:HMMER | 0.002192982 | -0.081973836 | 0.086359801 | 1 |
| dbCAN_2:Hotpep | dbCAN_2:DIAMOND | 0.002840434 | -0.081326384 | 0.087007253 | 1 |
| dbCAN_3:DIAMOND | dbCAN_3 | 0.001989348 | -0.08217747 | 0.086156167 | 1 |
| dbCAN_4:DIAMOND | dbCAN_4 | -0.002767335 | -0.086934154 | 0.081399484 | 1 |
| dbCAN_4:dbCANsub | dbCAN_4 | -0.002793442 | -0.086960261 | 0.081373377 | 1 |
| dbCAN_4:dbCANsub | dbCAN_4:DIAMOND | -2.61E-05 | -0.084192926 | 0.084140712 | 1 |

### 6.5 Auxiliary Activities

**SI table 44: Output of Tukey HSD test for statistically significant differences between the mean F1-score of tools classifying AA CAZyme domains (overleaf)**

| Classifier 1 | Classifier 2 | Mean Difference | Lower 95% Confidence Interval | Upper 95% Confidence Interval | Adjusted P-value |
|---|---|---|---|---|---|
| dbCAN_4 | dbCAN_3:eCAMI | 0.145848135 | 0.048717301 | 0.242978969 | 5.81E-05 |
| dbCAN_3:eCAMI | dbCAN_3 | -0.143028671 | -0.240159506 | -0.045897837 | 9.30E-05 |
| dbCAN_4:HMMER | dbCAN_3:eCAMI | 0.142847524 | 0.04571669 | 0.239978359 | 9.58E-05 |
| dbCAN_3:eCAMI | dbCAN_3:HMMER | -0.142847524 | -0.239978359 | -0.04571669 | 9.58E-05 |
| dbCAN_4:dbCANsub | dbCAN_3:eCAMI | 0.142388641 | 0.045257807 | 0.239519476 | 0.000103343 |
| dbCAN_3:eCAMI | dbCAN_3:DIAMOND | -0.140683072 | -0.237813906 | -0.043552238 | 0.00013662 |
| dbCAN_4:DIAMOND | dbCAN_3:eCAMI | 0.130042691 | 0.032911857 | 0.227173526 | 0.000722987 |
| dbCAN_3:eCAMI | dbCAN_2:HMMER | -0.116695045 | -0.21382588 | -0.019564211 | 0.004814228 |
| dbCAN_3:eCAMI | dbCAN_2 | -0.109818902 | -0.206949736 | -0.012688067 | 0.011671949 |
| CUPP | dbCAN_3:eCAMI | 0.100465436 | 0.003334602 | 0.19759627 | 0.03493286 |
| dbCAN_3:eCAMI | dbCAN_2:DIAMOND | -0.076033614 | -0.173164448 | 0.021097221 | 0.310300349 |
| dbCAN_4 | dbCAN_2:Hotpep | 0.073156807 | -0.023974027 | 0.170287641 | 0.372910083 |
| dbCAN_3:eCAMI | dbCAN_2:Hotpep | -0.072691328 | -0.169822162 | 0.024439507 | 0.383561592 |
| dbCAN_3 | dbCAN_2:Hotpep | 0.070337344 | -0.026793491 | 0.167468178 | 0.439275205 |
| dbCAN_4:HMMER | dbCAN_2:Hotpep | 0.070156197 | -0.026974638 | 0.167287031 | 0.443674197 |
| dbCAN_3:HMMER | dbCAN_2:Hotpep | 0.070156197 | -0.026974638 | 0.167287031 | 0.443674197 |
| dbCAN_4 | dbCAN_2:DIAMOND | 0.069814521 | -0.027316313 | 0.166945356 | 0.452008345 |
| dbCAN_4:dbCANsub | dbCAN_2:Hotpep | 0.069697313 | -0.027433521 | 0.166828148 | 0.454877917 |
| dbCAN_3:DIAMOND | dbCAN_2:Hotpep | 0.067991744 | -0.02913909 | 0.165122579 | 0.497153875 |
| dbCAN_3 | dbCAN_2:DIAMOND | 0.066995058 | -0.030135776 | 0.164125892 | 0.522194501 |
| dbCAN_4:HMMER | dbCAN_2:DIAMOND | 0.066813911 | -0.030316924 | 0.163944745 | 0.526762586 |
| dbCAN_3:HMMER | dbCAN_2:DIAMOND | 0.066813911 | -0.030316924 | 0.163944745 | 0.526762586 |
| dbCAN_4:dbCANsub | dbCAN_2:DIAMOND | 0.066355028 | -0.030775807 | 0.163485862 | 0.538350013 |
| dbCAN_3:DIAMOND | dbCAN_2:DIAMOND | 0.064649458 | -0.032481376 | 0.161780293 | 0.5814673 |
| dbCAN_4:DIAMOND | dbCAN_2:Hotpep | 0.057351364 | -0.039779471 | 0.154482198 | 0.756086028 |
| dbCAN_4:DIAMOND | dbCAN_2:DIAMOND | 0.054009078 | -0.043121756 | 0.151139912 | 0.823642207 |
| CUPP | dbCAN_4 | -0.045382699 | -0.142513533 | 0.051748135 | 0.942800769 |
| dbCAN_2:Hotpep | dbCAN_2:HMMER | -0.044003717 | -0.141134552 | 0.053127117 | 0.954375524 |
| CUPP | dbCAN_3 | -0.042563236 | -0.13969407 | 0.054567599 | 0.964538733 |
| CUPP | dbCAN_4:HMMER | -0.042382089 | -0.139512923 | 0.054748746 | 0.965685901 |
| CUPP | dbCAN_3:HMMER | -0.042382089 | -0.139512923 | 0.054748746 | 0.965685901 |
| CUPP | dbCAN_4:dbCANsub | -0.041923205 | -0.13905404 | 0.055207629 | 0.968467383 |
| dbCAN_2:DIAMOND | dbCAN_2:HMMER | -0.040661432 | -0.137792266 | 0.056469403 | 0.975239917 |
| CUPP | dbCAN_3:DIAMOND | -0.040217636 | -0.13734847 | 0.056913198 | 0.977334948 |
| dbCAN_2:Hotpep | dbCAN_2 | -0.037127574 | -0.134258408 | 0.060003261 | 0.98837662 |
| dbCAN_4 | dbCAN_2 | 0.036029233 | -0.061101601 | 0.133160068 | 0.991046961 |
| dbCAN_2:DIAMOND | dbCAN_2 | -0.033785288 | -0.130916122 | 0.063345546 | 0.994967582 |
| dbCAN_3 | dbCAN_2 | 0.03320977 | -0.063921065 | 0.130340604 | 0.995701576 |
| dbCAN_4:HMMER | dbCAN_2 | 0.033028623 | -0.064102212 | 0.130159457 | 0.995913267 |
| dbCAN_3:HMMER | dbCAN_2 | 0.033028623 | -0.064102212 | 0.130159457 | 0.995913267 |
| dbCAN_4:dbCANsub | dbCAN_2 | 0.03256974 | -0.064561095 | 0.129700574 | 0.996410964 |
| dbCAN_3:DIAMOND | dbCAN_2 | 0.03086417 | -0.066266664 | 0.127995005 | 0.997840458 |
| CUPP | dbCAN_4:DIAMOND | -0.029577256 | -0.12670809 | 0.067553579 | 0.998568959 |
| dbCAN_4 | dbCAN_2:HMMER | 0.02915309 | -0.067977745 | 0.126283924 | 0.998757538 |
| CUPP | dbCAN_2:Hotpep | 0.027774108 | -0.069356726 | 0.124904942 | 0.999231273 |
| dbCAN_3 | dbCAN_2:HMMER | 0.026333626 | -0.070797208 | 0.123464461 | 0.999551039 |
| dbCAN_3:HMMER | dbCAN_2:HMMER | 0.026152479 | -0.070978355 | 0.123283314 | 0.999581572 |
| dbCAN_4:HMMER | dbCAN_2:HMMER | 0.026152479 | -0.070978355 | 0.123283314 | 0.999581572 |
| dbCAN_4:dbCANsub | dbCAN_2:HMMER | 0.025693596 | -0.071437238 | 0.12282443 | 0.999650973 |
| CUPP | dbCAN_2:DIAMOND | 0.024431822 | -0.072699012 | 0.121562657 | 0.999792779 |
| dbCAN_3:DIAMOND | dbCAN_2:HMMER | 0.023988027 | -0.073142808 | 0.121118861 | 0.999828931 |
| dbCAN_4:DIAMOND | dbCAN_2 | 0.02022379 | -0.076907045 | 0.117354624 | 0.999972497 |
| CUPP | dbCAN_2:HMMER | -0.016229609 | -0.113360444 | 0.080901225 | 0.999997614 |
| dbCAN_4:DIAMOND | dbCAN_4 | -0.015805443 | -0.112936278 | 0.081325391 | 0.999998232 |
| dbCAN_4:DIAMOND | dbCAN_2:HMMER | 0.013347646 | -0.083783188 | 0.110478481 | 0.999999744 |
| dbCAN_4:DIAMOND | dbCAN_3 | -0.01298598 | -0.110116814 | 0.084144854 | 0.999999813 |
| dbCAN_4:DIAMOND | dbCAN_3:HMMER | -0.012804833 | -0.109935667 | 0.084326001 | 0.999999841 |
| dbCAN_4:DIAMOND | dbCAN_4:HMMER | -0.012804833 | -0.109935667 | 0.084326001 | 0.999999841 |
| dbCAN_4:dbCANsub | dbCAN_4:DIAMOND | 0.01234595 | -0.084784885 | 0.109476784 | 0.999999896 |
| dbCAN_4:DIAMOND | dbCAN_3:DIAMOND | -0.01064038 | -0.107771215 | 0.086490454 | 0.999999982 |
| CUPP | dbCAN_2 | -0.009353466 | -0.1064843 | 0.087777369 | 0.999999996 |
| dbCAN_2:HMMER | dbCAN_2 | 0.006876144 | -0.090254691 | 0.104006978 | 1 |
| dbCAN_4 | dbCAN_3:DIAMOND | 0.005165063 | -0.091965771 | 0.102295897 | 1 |
| dbCAN_2:Hotpep | dbCAN_2:DIAMOND | -0.003342286 | -0.10047312 | 0.093788549 | 1 |
| dbCAN_3:HMMER | dbCAN_3 | -0.000181147 | -0.097311981 | 0.096949687 | 1 |
| dbCAN_3:DIAMOND | dbCAN_3 | -0.002345599 | -0.099476434 | 0.094785235 | 1 |
| dbCAN_4 | dbCAN_3 | 0.002819463 | -0.094311371 | 0.099950298 | 1 |
| dbCAN_4:HMMER | dbCAN_3 | -0.000181147 | -0.097311981 | 0.096949687 | 1 |
| dbCAN_4:dbCANsub | dbCAN_3 | -0.00064003 | -0.097770865 | 0.096490804 | 1 |
| dbCAN_3:DIAMOND | dbCAN_3:HMMER | -0.002164452 | -0.099295287 | 0.094966382 | 1 |
| dbCAN_4 | dbCAN_3:HMMER | 0.003000611 | -0.094130224 | 0.100131445 | 1 |
| dbCAN_4:HMMER | dbCAN_3:HMMER | 8.88E-16 | -0.097130834 | 0.097130834 | 1 |
| dbCAN_4:dbCANsub | dbCAN_3:HMMER | -0.000458883 | -0.097589718 | 0.096671951 | 1 |
| dbCAN_4:HMMER | dbCAN_3:DIAMOND | 0.002164452 | -0.094966382 | 0.099295287 | 1 |
| dbCAN_4:dbCANsub | dbCAN_3:DIAMOND | 0.001705569 | -0.095425265 | 0.098836404 | 1 |
| dbCAN_4:HMMER | dbCAN_4 | -0.003000611 | -0.100131445 | 0.094130224 | 1 |
| dbCAN_4:dbCANsub | dbCAN_4 | -0.003459494 | -0.100590328 | 0.093671341 | 1 |
| dbCAN_4:dbCANsub | dbCAN_4:HMMER | -0.000458883 | -0.097589718 | 0.096671951 | 1 |

**SI table 45: Output of Tukey HSD test for statistically significant differences between the mean sensitivity of tools classifying AA CAZyme domains (overleaf)**

| Classifier 1 | Classifier 2 | Mean Difference | Lower 95% Confidence Interval | Upper 95% Confidence Interval | Adjusted P-value |
|---|---|---|---|---|---|
| dbCAN_4 | dbCAN_3:eCAMI | 0.211614254 | 0.128762009 | 0.294466498 | 3.54E-10 |
| dbCAN_4:dbCANsub | dbCAN_3:eCAMI | 0.207116148 | 0.124263903 | 0.289968392 | 3.54E-10 |
| dbCAN_3:eCAMI | dbCAN_3:DIAMOND | -0.206291452 | -0.289143696 | -0.123439207 | 3.54E-10 |
| dbCAN_3:eCAMI | dbCAN_3:HMMER | -0.205917404 | -0.288769649 | -0.12306516 | 3.54E-10 |
| dbCAN_4:HMMER | dbCAN_3:eCAMI | 0.205917404 | 0.12306516 | 0.288769649 | 3.54E-10 |
| dbCAN_3:eCAMI | dbCAN_3 | -0.201462658 | -0.284314902 | -0.118610413 | 3.54E-10 |
| dbCAN_4:DIAMOND | dbCAN_3:eCAMI | 0.187919199 | 0.105066954 | 0.270771443 | 3.67E-10 |
| dbCAN_3:eCAMI | dbCAN_2:HMMER | -0.155417615 | -0.238269859 | -0.07256537 | 6.43E-08 |
| dbCAN_3:eCAMI | dbCAN_2 | -0.136912451 | -0.219764696 | -0.054060206 | 4.38E-06 |
| dbCAN_4 | dbCAN_2:DIAMOND | 0.124369612 | 0.041517367 | 0.207221856 | 5.85E-05 |
| CUPP | dbCAN_3:eCAMI | 0.122659054 | 0.039806809 | 0.205511298 | 8.19E-05 |
| dbCAN_4 | dbCAN_2:Hotpep | 0.119968689 | 0.037116444 | 0.202820933 | 0.000137494 |
| dbCAN_4:dbCANsub | dbCAN_2:DIAMOND | 0.119871506 | 0.037019261 | 0.20272375 | 0.000140065 |
| dbCAN_3:DIAMOND | dbCAN_2:DIAMOND | 0.11904681 | 0.036194565 | 0.201899054 | 0.000163813 |
| dbCAN_3:HMMER | dbCAN_2:DIAMOND | 0.118672762 | 0.035820518 | 0.201525007 | 0.000175812 |
| dbCAN_4:HMMER | dbCAN_2:DIAMOND | 0.118672762 | 0.035820518 | 0.201525007 | 0.000175812 |
| dbCAN_4:dbCANsub | dbCAN_2:Hotpep | 0.115470583 | 0.032618338 | 0.198322827 | 0.000319115 |
| dbCAN_3:DIAMOND | dbCAN_2:Hotpep | 0.114645887 | 0.031793642 | 0.197498131 | 0.00037109 |
| dbCAN_3:HMMER | dbCAN_2:Hotpep | 0.114271839 | 0.031419595 | 0.197124084 | 0.000397233 |
| dbCAN_4:HMMER | dbCAN_2:Hotpep | 0.114271839 | 0.031419595 | 0.197124084 | 0.000397233 |
| dbCAN_3 | dbCAN_2:DIAMOND | 0.114218016 | 0.031365771 | 0.19707026 | 0.000401136 |
| dbCAN_3 | dbCAN_2:Hotpep | 0.109817093 | 0.026964848 | 0.192669337 | 0.000878133 |
| dbCAN_4:DIAMOND | dbCAN_2:DIAMOND | 0.100674557 | 0.017822312 | 0.183526802 | 0.004026719 |
| dbCAN_4:DIAMOND | dbCAN_2:Hotpep | 0.096273634 | 0.013421389 | 0.179125878 | 0.007947441 |
| dbCAN_3:eCAMI | dbCAN_2:Hotpep | -0.091645565 | -0.17449781 | -0.00879332 | 0.015615318 |
| CUPP | dbCAN_4 | -0.0889552 | -0.171807444 | -0.006102955 | 0.022678396 |
| dbCAN_3:eCAMI | dbCAN_2:DIAMOND | -0.087244642 | -0.170096887 | -0.004392397 | 0.028530398 |
| CUPP | dbCAN_4:dbCANsub | -0.084457094 | -0.167309338 | -0.001604849 | 0.040929858 |
| CUPP | dbCAN_3:DIAMOND | -0.083632398 | -0.166484642 | -0.000780153 | 0.045396004 |
| CUPP | dbCAN_4:HMMER | -0.08325835 | -0.166110595 | -0.000406106 | 0.047555909 |
| CUPP | dbCAN_3:HMMER | -0.08325835 | -0.166110595 | -0.000406106 | 0.047555909 |
| CUPP | dbCAN_3 | -0.078803604 | -0.161655848 | 0.004048641 | 0.08075257 |
| dbCAN_4 | dbCAN_2 | 0.074701803 | -0.008150442 | 0.157554047 | 0.12622594 |
| dbCAN_4:dbCANsub | dbCAN_2 | 0.070203697 | -0.012648548 | 0.153055941 | 0.196432946 |
| dbCAN_3:DIAMOND | dbCAN_2 | 0.069379001 | -0.013473244 | 0.152231245 | 0.211831191 |
| dbCAN_4:HMMER | dbCAN_2 | 0.069004953 | -0.013847291 | 0.151857198 | 0.219079033 |
| dbCAN_3:HMMER | dbCAN_2 | 0.069004953 | -0.013847291 | 0.151857198 | 0.219079033 |
| dbCAN_2:DIAMOND | dbCAN_2:HMMER | -0.068172973 | -0.151025217 | 0.014679272 | 0.235789791 |
| CUPP | dbCAN_4:DIAMOND | -0.065260145 | -0.14811239 | 0.0175921 | 0.300577914 |
| dbCAN_3 | dbCAN_2 | 0.064550207 | -0.018302038 | 0.147402451 | 0.317787702 |
| dbCAN_2:Hotpep | dbCAN_2:HMMER | -0.06377205 | -0.146624294 | 0.019080195 | 0.3372472 |
| dbCAN_4 | dbCAN_2:HMMER | 0.056196639 | -0.026655606 | 0.139048884 | 0.550323504 |
| dbCAN_4:dbCANsub | dbCAN_2:HMMER | 0.051698533 | -0.031153712 | 0.134550778 | 0.681440246 |
| dbCAN_4:DIAMOND | dbCAN_2 | 0.051006748 | -0.031845497 | 0.133858992 | 0.700681683 |
| dbCAN_3:DIAMOND | dbCAN_2:HMMER | 0.050873837 | -0.031978408 | 0.133726081 | 0.704334289 |
| dbCAN_4:HMMER | dbCAN_2:HMMER | 0.05049979 | -0.032352455 | 0.133352034 | 0.714531326 |
| dbCAN_3:HMMER | dbCAN_2:HMMER | 0.05049979 | -0.032352455 | 0.133352034 | 0.714531326 |
| dbCAN_2:DIAMOND | dbCAN_2 | -0.049667809 | -0.132520054 | 0.033184436 | 0.736742735 |
| dbCAN_3 | dbCAN_2:HMMER | 0.046045043 | -0.036807202 | 0.128897288 | 0.824175398 |
| dbCAN_2:Hotpep | dbCAN_2 | -0.045266886 | -0.12811913 | 0.037585359 | 0.840655944 |
| CUPP | dbCAN_2:DIAMOND | 0.035414412 | -0.047437833 | 0.118266657 | 0.970780869 |
| CUPP | dbCAN_2:HMMER | -0.032758561 | -0.115610805 | 0.050093684 | 0.984507477 |
| dbCAN_4:DIAMOND | dbCAN_2:HMMER | 0.032501584 | -0.050350661 | 0.115353829 | 0.985501458 |
| CUPP | dbCAN_2:Hotpep | 0.031013489 | -0.051838756 | 0.113865733 | 0.990305306 |
| dbCAN_4:DIAMOND | dbCAN_4 | -0.023695055 | -0.1065473 | 0.05915719 | 0.999230025 |
| dbCAN_4:dbCANsub | dbCAN_4:DIAMOND | 0.019196949 | -0.063655296 | 0.102049193 | 0.999912867 |
| dbCAN_2:HMMER | dbCAN_2 | 0.018505164 | -0.064347081 | 0.101357408 | 0.999941177 |
| dbCAN_4:DIAMOND | dbCAN_3:DIAMOND | -0.018372253 | -0.101224497 | 0.064479992 | 0.999945566 |
| dbCAN_4:DIAMOND | dbCAN_3:HMMER | -0.017998205 | -0.10085045 | 0.064854039 | 0.999956402 |
| dbCAN_4:DIAMOND | dbCAN_4:HMMER | -0.017998205 | -0.10085045 | 0.064854039 | 0.999956402 |
| CUPP | dbCAN_2 | -0.014253397 | -0.097105642 | 0.068598848 | 0.999996686 |
| dbCAN_4:DIAMOND | dbCAN_3 | -0.013543459 | -0.096395703 | 0.069308786 | 0.999998138 |
| dbCAN_4 | dbCAN_3 | 0.010151596 | -0.072700648 | 0.093003841 | 0.999999932 |
| dbCAN_4 | dbCAN_3:HMMER | 0.005696849 | -0.077155395 | 0.088549094 | 1 |
| dbCAN_4:HMMER | dbCAN_4 | -0.005696849 | -0.088549094 | 0.077155395 | 1 |
| dbCAN_4:dbCANsub | dbCAN_3 | 0.00565349 | -0.077198755 | 0.088505735 | 1 |
| dbCAN_4 | dbCAN_3:DIAMOND | 0.005322802 | -0.077529442 | 0.088175047 | 1 |
| dbCAN_3:DIAMOND | dbCAN_3 | 0.004828794 | -0.078023451 | 0.087681038 | 1 |
| dbCAN_4:dbCANsub | dbCAN_4 | -0.004498106 | -0.087350351 | 0.078354139 | 1 |
| dbCAN_3:HMMER | dbCAN_3 | 0.004454747 | -0.078397498 | 0.087306991 | 1 |
| dbCAN_4:HMMER | dbCAN_3 | 0.004454747 | -0.078397498 | 0.087306991 | 1 |
| dbCAN_2:Hotpep | dbCAN_2:DIAMOND | 0.004400923 | -0.078451321 | 0.087253168 | 1 |
| dbCAN_3:DIAMOND | dbCAN_3:HMMER | 0.000374047 | -0.082478197 | 0.083226292 | 1 |
| dbCAN_4:HMMER | dbCAN_3:HMMER | 7.77E-16 | -0.082852245 | 0.082852245 | 1 |
| dbCAN_4:dbCANsub | dbCAN_3:HMMER | 0.001198743 | -0.081653501 | 0.084050988 | 1 |
| dbCAN_4:HMMER | dbCAN_3:DIAMOND | -0.000374047 | -0.083226292 | 0.082478197 | 1 |
| dbCAN_4:dbCANsub | dbCAN_3:DIAMOND | 0.000824696 | -0.082027548 | 0.083676941 | 1 |
| dbCAN_4:dbCANsub | dbCAN_4:HMMER | 0.001198743 | -0.081653501 | 0.084050988 | 1 |

### 6.6 Carbohydrate Binding Modules

**SI table 46: Output of Tukey HSD test for statistically significant differences between the mean specificity of tools classifying CBM CAZyme domains (overleaf)**

| Classifier 1 | Classifier 2 | Mean Difference | Lower 95% Confidence Interval | Upper 95% Confidence Interval | Adjusted P-value |
|---|---|---|---|---|---|
| dbCAN_2:Hotpep | dbCAN_2 | -0.090754579 | -0.102742275 | -0.078766884 | 2.86E-13 |
| dbCAN_3:eCAMI | dbCAN_2 | -0.04517558 | -0.057163275 | -0.033187884 | 2.86E-13 |
| dbCAN_2:Hotpep | dbCAN_2:HMMER | -0.094603713 | -0.106591409 | -0.082616018 | 2.86E-13 |
| dbCAN_3:eCAMI | dbCAN_2:HMMER | -0.049024714 | -0.061012409 | -0.037037018 | 2.86E-13 |
| dbCAN_2:Hotpep | dbCAN_2:DIAMOND | -0.089950771 | -0.101938466 | -0.077963075 | 2.86E-13 |
| dbCAN_3:eCAMI | dbCAN_2:DIAMOND | -0.044371771 | -0.056359467 | -0.032384076 | 2.86E-13 |
| dbCAN_3 | dbCAN_2:Hotpep | 0.092896328 | 0.080908632 | 0.104884023 | 2.86E-13 |
| dbCAN_3:HMMER | dbCAN_2:Hotpep | 0.094593607 | 0.082605911 | 0.106581302 | 2.86E-13 |
| dbCAN_3:DIAMOND | dbCAN_2:Hotpep | 0.091613333 | 0.079625637 | 0.103601028 | 2.86E-13 |
| dbCAN_3:eCAMI | dbCAN_2:Hotpep | 0.045578999 | 0.033591304 | 0.057566695 | 2.86E-13 |
| dbCAN_4 | dbCAN_2:Hotpep | 0.093671037 | 0.081683342 | 0.105658733 | 2.86E-13 |
| dbCAN_4:HMMER | dbCAN_2:Hotpep | 0.094364249 | 0.082376554 | 0.106351944 | 2.86E-13 |
| dbCAN_4:DIAMOND | dbCAN_2:Hotpep | 0.092358026 | 0.080370331 | 0.104345721 | 2.86E-13 |
| dbCAN_4:dbCANsub | dbCAN_2:Hotpep | 0.09122322 | 0.079235525 | 0.103210916 | 2.86E-13 |
| CUPP | dbCAN_2:Hotpep | 0.09855361 | 0.086565914 | 0.110541305 | 2.86E-13 |
| dbCAN_3:eCAMI | dbCAN_3 | -0.047317328 | -0.059305024 | -0.035329633 | 2.86E-13 |
| dbCAN_3:eCAMI | dbCAN_3:HMMER | -0.049014607 | -0.061002303 | -0.037026912 | 2.86E-13 |
| dbCAN_3:eCAMI | dbCAN_3:DIAMOND | -0.046034333 | -0.058022029 | -0.034046638 | 2.86E-13 |
| dbCAN_4 | dbCAN_3:eCAMI | 0.048092038 | 0.036104343 | 0.060079734 | 2.86E-13 |
| dbCAN_4:HMMER | dbCAN_3:eCAMI | 0.04878525 | 0.036797554 | 0.060772945 | 2.86E-13 |
| dbCAN_4:DIAMOND | dbCAN_3:eCAMI | 0.046779027 | 0.034791331 | 0.058766722 | 2.86E-13 |
| dbCAN_4:dbCANsub | dbCAN_3:eCAMI | 0.045644221 | 0.033656525 | 0.057631916 | 2.86E-13 |
| CUPP | dbCAN_3:eCAMI | 0.05297461 | 0.040986915 | 0.064962306 | 2.86E-13 |
| CUPP | dbCAN_2:DIAMOND | 0.008602839 | -0.003384857 | 0.020590534 | 0.45687775 |
| CUPP | dbCAN_2 | 0.00779903 | -0.004188665 | 0.019786726 | 0.620558321 |
| CUPP | dbCAN_4:dbCANsub | 0.007330389 | -0.004657306 | 0.019318085 | 0.712468856 |
| CUPP | dbCAN_3:DIAMOND | 0.006940277 | -0.005047419 | 0.018927972 | 0.782004989 |
| CUPP | dbCAN_4:DIAMOND | 0.006195584 | -0.005792112 | 0.018183279 | 0.888036638 |
| CUPP | dbCAN_3 | 0.005657282 | -0.006330414 | 0.017644977 | 0.939548681 |
| CUPP | dbCAN_4 | 0.004882572 | -0.007105123 | 0.016870268 | 0.980583861 |
| dbCAN_2:DIAMOND | dbCAN_2:HMMER | -0.004652942 | -0.016640638 | 0.007334753 | 0.987036379 |
| dbCAN_3:HMMER | dbCAN_2:DIAMOND | 0.004642836 | -0.00734486 | 0.016630531 | 0.987274982 |
| dbCAN_4:HMMER | dbCAN_2:DIAMOND | 0.004413478 | -0.007574217 | 0.016401174 | 0.991811264 |
| CUPP | dbCAN_4:HMMER | 0.004189361 | -0.007798335 | 0.016177056 | 0.994874751 |
| CUPP | dbCAN_3:HMMER | 0.003960003 | -0.008027693 | 0.015947698 | 0.996959324 |
| CUPP | dbCAN_2:HMMER | 0.003949896 | -0.008037799 | 0.015937592 | 0.997031621 |
| dbCAN_2:HMMER | dbCAN_2 | 0.003849134 | -0.008138561 | 0.015836829 | 0.997676772 |
| dbCAN_3:HMMER | dbCAN_2 | 0.003839028 | -0.008148668 | 0.015826723 | 0.997734368 |
| dbCAN_4 | dbCAN_2:DIAMOND | 0.003720267 | -0.008267429 | 0.015707962 | 0.998325649 |
| dbCAN_4:HMMER | dbCAN_2 | 0.00360967 | -0.008378026 | 0.015597365 | 0.998752682 |
| dbCAN_4:dbCANsub | dbCAN_2:HMMER | -0.003380493 | -0.015368188 | 0.008607203 | 0.999350447 |
| dbCAN_4:dbCANsub | dbCAN_3:HMMER | -0.003370386 | -0.015358082 | 0.008617309 | 0.99936975 |
| dbCAN_4:dbCANsub | dbCAN_4:HMMER | -0.003141029 | -0.015128724 | 0.008846667 | 0.999692952 |
| dbCAN_3:DIAMOND | dbCAN_2:HMMER | -0.00299038 | -0.014978076 | 0.008997315 | 0.999815802 |
| dbCAN_3:DIAMOND | dbCAN_3:HMMER | -0.002980274 | -0.014967969 | 0.009007422 | 0.99982222 |
| dbCAN_3 | dbCAN_2:DIAMOND | 0.002945557 | -0.009042139 | 0.014933252 | 0.999842795 |
| dbCAN_4 | dbCAN_2 | 0.002916458 | -0.009071237 | 0.014904154 | 0.999858397 |
| dbCAN_4:HMMER | dbCAN_3:DIAMOND | 0.002750916 | -0.009236779 | 0.014738612 | 0.999923872 |
| dbCAN_4:dbCANsub | dbCAN_4 | -0.002447817 | -0.014435513 | 0.009539878 | 0.999978504 |
| dbCAN_4:DIAMOND | dbCAN_2:DIAMOND | 0.002407255 | -0.00958044 | 0.014394951 | 0.999982106 |
| dbCAN_4:DIAMOND | dbCAN_2:HMMER | -0.002245687 | -0.014233383 | 0.009742008 | 0.9999917 |
| dbCAN_4:DIAMOND | dbCAN_3:HMMER | -0.002235581 | -0.014223276 | 0.009752115 | 0.999992106 |
| dbCAN_3 | dbCAN_2 | 0.002141749 | -0.009845947 | 0.014129444 | 0.999995109 |
| dbCAN_4 | dbCAN_3:DIAMOND | 0.002057705 | -0.009929991 | 0.0140454 | 0.99999688 |
| dbCAN_4:DIAMOND | dbCAN_4:HMMER | -0.002006223 | -0.013993918 | 0.009981472 | 0.999997655 |
| dbCAN_3 | dbCAN_2:HMMER | -0.001707385 | -0.013695081 | 0.01028031 | 0.999999627 |
| dbCAN_3:HMMER | dbCAN_3 | 0.001697279 | -0.010290416 | 0.013684974 | 0.999999651 |
| dbCAN_4:dbCANsub | dbCAN_3 | -0.001673107 | -0.013660803 | 0.010314588 | 0.999999704 |
| dbCAN_3:DIAMOND | dbCAN_2:DIAMOND | 0.001662562 | -0.010325133 | 0.013650257 | 0.999999725 |
| dbCAN_4:DIAMOND | dbCAN_2 | 0.001603447 | -0.010384249 | 0.013591142 | 0.999999819 |
| dbCAN_4:HMMER | dbCAN_3 | 0.001467921 | -0.010519774 | 0.013455617 | 0.999999935 |
| dbCAN_4:DIAMOND | dbCAN_4 | -0.001313011 | -0.013300707 | 0.010674684 | 0.999999982 |
| dbCAN_3:DIAMOND | dbCAN_3 | -0.001282995 | -0.01327069 | 0.010704701 | 0.999999986 |
| dbCAN_4:dbCANsub | dbCAN_2:DIAMOND | 0.001272449 | -0.010715246 | 0.013260145 | 0.999999988 |
| dbCAN_4:dbCANsub | dbCAN_4:DIAMOND | -0.001134806 | -0.013122501 | 0.01085289 | 0.999999997 |
| dbCAN_4 | dbCAN_2:HMMER | -0.000932676 | -0.012920371 | 0.01105502 | 1 |
| dbCAN_4 | dbCAN_3:HMMER | -0.000922569 | -0.012910265 | 0.011065126 | 1 |
| dbCAN_3:DIAMOND | dbCAN_2 | 0.000858754 | -0.011128942 | 0.012846449 | 1 |
| dbCAN_2:DIAMOND | dbCAN_2 | -0.000803808 | -0.012791504 | 0.011183887 | 1 |
| dbCAN_4 | dbCAN_3 | 0.00077471 | -0.011212986 | 0.012762405 | 1 |
| dbCAN_4:DIAMOND | dbCAN_3:DIAMOND | 0.000744693 | -0.011243002 | 0.012732389 | 1 |
| dbCAN_4:HMMER | dbCAN_4 | 0.000693212 | -0.011294484 | 0.012680907 | 1 |
| dbCAN_4:DIAMOND | dbCAN_3 | -0.000538302 | -0.012525997 | 0.011449394 | 1 |
| dbCAN_4:dbCANsub | dbCAN_2 | 0.000468641 | -0.011519054 | 0.012456337 | 1 |
| dbCAN_3:HMMER | dbCAN_2:HMMER | -1.01E-05 | -0.011997802 | 0.011977589 | 1 |
| dbCAN_4:HMMER | dbCAN_2:HMMER | -0.000239464 | -0.01222716 | 0.011748231 | 1 |
| dbCAN_4:HMMER | dbCAN_3:HMMER | -0.000229358 | -0.012217053 | 0.011758338 | 1 |
| dbCAN_4:dbCANsub | dbCAN_3:DIAMOND | -0.000390113 | -0.012377808 | 0.011597583 | 1 |

**SI table 47: Output of Tukey HSD test for statistically significant differences between the mean precision of tools classifying CBM CAZyme domains (overleaf)**

| Classifier 1 | Classifier 2 | Mean Difference | Lower 95% Confidence Interval | Upper 95% Confidence Interval | Adjusted P-value |
|---|---|---|---|---|---|
| dbCAN_2:Hotpep | dbCAN_2 | -0.450661933 | -0.535567674 | -0.365756191 | 2.86E-13 |
| CUPP | dbCAN_2 | -0.920284789 | -1.005190531 | -0.835379048 | 2.86E-13 |
| dbCAN_2:Hotpep | dbCAN_2:HMMER | -0.426880432 | -0.511786173 | -0.34197469 | 2.86E-13 |
| CUPP | dbCAN_2:HMMER | -0.896503288 | -0.98140903 | -0.811597547 | 2.86E-13 |
| dbCAN_2:Hotpep | dbCAN_2:DIAMOND | -0.45230696 | -0.537212701 | -0.367401219 | 2.86E-13 |
| CUPP | dbCAN_2:DIAMOND | -0.921929817 | -1.006835558 | -0.837024075 | 2.86E-13 |
| dbCAN_3 | dbCAN_2:Hotpep | 0.47343433 | 0.388528589 | 0.558340071 | 2.86E-13 |
| dbCAN_3:HMMER | dbCAN_2:Hotpep | 0.426750261 | 0.34184452 | 0.511656002 | 2.86E-13 |
| dbCAN_3:DIAMOND | dbCAN_2:Hotpep | 0.481573872 | 0.396668131 | 0.566479613 | 2.86E-13 |
| dbCAN_4 | dbCAN_2:Hotpep | 0.486625171 | 0.40171943 | 0.571530913 | 2.86E-13 |
| dbCAN_4:HMMER | dbCAN_2:Hotpep | 0.448772718 | 0.363866976 | 0.533678459 | 2.86E-13 |
| dbCAN_4:DIAMOND | dbCAN_2:Hotpep | 0.487484714 | 0.402578973 | 0.572390456 | 2.86E-13 |
| dbCAN_4:dbCANsub | dbCAN_2:Hotpep | 0.467970775 | 0.383065034 | 0.552876517 | 2.86E-13 |
| CUPP | dbCAN_2:Hotpep | -0.469622857 | -0.554528598 | -0.384717115 | 2.86E-13 |
| CUPP | dbCAN_3 | -0.943057187 | -1.027962928 | -0.858151445 | 2.86E-13 |
| CUPP | dbCAN_3:HMMER | -0.896373118 | -0.981278859 | -0.811467376 | 2.86E-13 |
| CUPP | dbCAN_3:DIAMOND | -0.951196729 | -1.03610247 | -0.866290987 | 2.86E-13 |
| CUPP | dbCAN_3:eCAMI | -0.661506581 | -0.746412322 | -0.576600839 | 2.86E-13 |
| CUPP | dbCAN_4 | -0.956248028 | -1.041153769 | -0.871342287 | 2.86E-13 |
| CUPP | dbCAN_4:HMMER | -0.918395574 | -1.003301316 | -0.833489833 | 2.86E-13 |
| CUPP | dbCAN_4:DIAMOND | -0.957107571 | -1.042013312 | -0.87220183 | 2.86E-13 |
| CUPP | dbCAN_4:dbCANsub | -0.937593632 | -1.022499373 | -0.852687891 | 2.86E-13 |
| dbCAN_4:DIAMOND | dbCAN_3:eCAMI | 0.29560099 | 0.210695249 | 0.380506732 | 2.93E-13 |
| dbCAN_4 | dbCAN_3:eCAMI | 0.294741447 | 0.209835706 | 0.379647189 | 2.94E-13 |
| dbCAN_3:eCAMI | dbCAN_3:DIAMOND | -0.289690148 | -0.374595889 | -0.204784407 | 3.12E-13 |
| dbCAN_3:eCAMI | dbCAN_3 | -0.281550606 | -0.366456347 | -0.196644865 | 3.65E-13 |
| dbCAN_4:dbCANsub | dbCAN_3:eCAMI | 0.276087051 | 0.19118131 | 0.360992793 | 3.96E-13 |
| dbCAN_3:eCAMI | dbCAN_2 | -0.258778209 | -0.34368395 | -0.173872467 | 4.08E-13 |
| dbCAN_3:eCAMI | dbCAN_2:DIAMOND | -0.260423236 | -0.345328977 | -0.175517495 | 4.10E-13 |
| dbCAN_4:HMMER | dbCAN_3:eCAMI | 0.256888994 | 0.171983252 | 0.341794735 | 4.11E-13 |
| dbCAN_3:eCAMI | dbCAN_2:HMMER | -0.234996708 | -0.319902449 | -0.150090966 | 4.19E-13 |
| dbCAN_3:eCAMI | dbCAN_3:HMMER | -0.234866537 | -0.319772278 | -0.149960796 | 4.20E-13 |
| dbCAN_3:eCAMI | dbCAN_2:Hotpep | 0.191883724 | 0.106977983 | 0.276789465 | 1.08E-11 |
| dbCAN_4:DIAMOND | dbCAN_3:HMMER | 0.060734453 | -0.024171288 | 0.145640195 | 0.462433123 |
| dbCAN_4:DIAMOND | dbCAN_2:HMMER | 0.060604283 | -0.024301459 | 0.145510024 | 0.466109621 |
| dbCAN_4 | dbCAN_3:HMMER | 0.05987491 | -0.025030831 | 0.144780652 | 0.486840717 |
| dbCAN_4 | dbCAN_2:HMMER | 0.05974474 | -0.025161002 | 0.144650481 | 0.490561488 |
| dbCAN_3:DIAMOND | dbCAN_3:HMMER | 0.054823611 | -0.03008213 | 0.139729352 | 0.632380244 |
| dbCAN_3:DIAMOND | dbCAN_2:HMMER | 0.05469344 | -0.030212301 | 0.139599182 | 0.636074728 |
| dbCAN_3:HMMER | dbCAN_3 | -0.046684069 | -0.13158981 | 0.038221672 | 0.836551474 |
| dbCAN_3 | dbCAN_2:HMMER | 0.046553898 | -0.038351843 | 0.13145964 | 0.839194869 |
| dbCAN_4:dbCANsub | dbCAN_3:HMMER | 0.041220514 | -0.043685227 | 0.126126256 | 0.926198498 |
| dbCAN_4:dbCANsub | dbCAN_2:HMMER | 0.041090344 | -0.043815398 | 0.125996085 | 0.927800876 |
| dbCAN_4:DIAMOND | dbCAN_4:HMMER | 0.038711997 | -0.046193745 | 0.123617738 | 0.952981761 |
| dbCAN_4:HMMER | dbCAN_4 | -0.037852454 | -0.122758195 | 0.047053288 | 0.960280518 |
| dbCAN_4:DIAMOND | dbCAN_2 | 0.036822782 | -0.04808296 | 0.121728523 | 0.967878876 |
| dbCAN_4 | dbCAN_2 | 0.035963239 | -0.048942503 | 0.12086898 | 0.97333581 |
| dbCAN_4:DIAMOND | dbCAN_2:DIAMOND | 0.035177754 | -0.049727987 | 0.120083496 | 0.977674546 |
| dbCAN_4 | dbCAN_2:DIAMOND | 0.034318211 | -0.05058753 | 0.119223953 | 0.981773277 |
| dbCAN_4:HMMER | dbCAN_3:DIAMOND | -0.032801154 | -0.117706896 | 0.052104587 | 0.98754675 |
| dbCAN_3:DIAMOND | dbCAN_2 | 0.030911939 | -0.053993802 | 0.115817681 | 0.992585225 |
| dbCAN_3:DIAMOND | dbCAN_2:DIAMOND | 0.029266912 | -0.055638829 | 0.114172653 | 0.995481731 |
| dbCAN_3:HMMER | dbCAN_2:DIAMOND | -0.025556699 | -0.11046244 | 0.059349042 | 0.998757301 |
| dbCAN_2:DIAMOND | dbCAN_2:HMMER | 0.025426528 | -0.059479213 | 0.11033227 | 0.998818142 |
| dbCAN_4:HMMER | dbCAN_3 | -0.024661612 | -0.109567354 | 0.060244129 | 0.999126584 |
| dbCAN_3:HMMER | dbCAN_2 | -0.023911672 | -0.108817413 | 0.06099407 | 0.999358999 |
| dbCAN_2:HMMER | dbCAN_2 | -0.023781501 | -0.108687242 | 0.06112424 | 0.999393334 |
| dbCAN_3 | dbCAN_2 | 0.022772397 | -0.062133344 | 0.107678139 | 0.999609673 |
| dbCAN_4:HMMER | dbCAN_3:HMMER | 0.022022457 | -0.062883285 | 0.106928198 | 0.99972353 |
| dbCAN_4:HMMER | dbCAN_2:HMMER | 0.021892286 | -0.063013455 | 0.106798027 | 0.999740002 |
| dbCAN_3 | dbCAN_2:DIAMOND | 0.02112737 | -0.063778371 | 0.106033111 | 0.999820548 |
| dbCAN_4:dbCANsub | dbCAN_4:DIAMOND | -0.019513939 | -0.10441968 | 0.065391802 | 0.999922613 |
| dbCAN_4:dbCANsub | dbCAN_4:HMMER | 0.019198058 | -0.065707684 | 0.104103799 | 0.999935026 |
| dbCAN_4:dbCANsub | dbCAN_4 | -0.018654396 | -0.103560137 | 0.066251345 | 0.999952312 |
| dbCAN_4:dbCANsub | dbCAN_2 | 0.017308843 | -0.067596899 | 0.102214584 | 0.999978887 |
| dbCAN_4:dbCANsub | dbCAN_2:DIAMOND | 0.015663815 | -0.069241926 | 0.100569557 | 0.999993002 |
| dbCAN_4:DIAMOND | dbCAN_3 | 0.014050384 | -0.070855357 | 0.098956126 | 0.999997935 |
| dbCAN_4:dbCANsub | dbCAN_3:DIAMOND | -0.013603097 | -0.098508838 | 0.071302645 | 0.999998569 |
| dbCAN_4 | dbCAN_3 | 0.013190841 | -0.0717149 | 0.098096583 | 0.999998992 |
| dbCAN_3:DIAMOND | dbCAN_3 | 0.008139542 | -0.076766199 | 0.093045283 | 0.999999996 |
| dbCAN_4:DIAMOND | dbCAN_3:DIAMOND | 0.005910842 | -0.078994899 | 0.090816584 | 1 |
| dbCAN_4:dbCANsub | dbCAN_3 | -0.005463555 | -0.090369296 | 0.079442187 | 1 |
| dbCAN_4 | dbCAN_3:DIAMOND | 0.005051299 | -0.079854442 | 0.089957041 | 1 |
| dbCAN_2:DIAMOND | dbCAN_2 | 0.001645027 | -0.083260714 | 0.086550769 | 1 |
| dbCAN_4:HMMER | dbCAN_2 | -0.001889215 | -0.086794956 | 0.083016526 | 1 |
| dbCAN_3:HMMER | dbCAN_2:HMMER | -0.000130171 | -0.085035912 | 0.084775571 | 1 |
| dbCAN_4:HMMER | dbCAN_2:DIAMOND | -0.003534242 | -0.088439984 | 0.081371499 | 1 |
| dbCAN_4:DIAMOND | dbCAN_4 | 0.000859543 | -0.084046198 | 0.085765284 | 1 |

**SI table 48: Output of Tukey HSD test for statistically significant differences between the mean F1-score of tools classifying CBM CAZyme domains (overleaf)**

| Classifier 1 | Classifier 2 | Mean Difference | Lower 95% Confidence Interval | Upper 95% Confidence Interval | Adjusted P-value |
|---|---|---|---|---|---|
| CUPP | dbCAN_2 | -0.797034509 | -0.885755905 | -0.708313114 | 2.86E-13 |
| dbCAN_3:DIAMOND | dbCAN_2:HMMER | 0.357201231 | 0.268479835 | 0.445922627 | 2.86E-13 |
| dbCAN_4:DIAMOND | dbCAN_2:HMMER | 0.365788976 | 0.27706758 | 0.454510372 | 2.86E-13 |
| CUPP | dbCAN_2:HMMER | -0.541375016 | -0.630096412 | -0.45265362 | 2.86E-13 |
| CUPP | dbCAN_2:DIAMOND | -0.832532715 | -0.921254111 | -0.743811319 | 2.86E-13 |
| dbCAN_3:DIAMOND | dbCAN_2:Hotpep | 0.347759862 | 0.259038466 | 0.436481258 | 2.86E-13 |
| dbCAN_4:DIAMOND | dbCAN_2:Hotpep | 0.356347607 | 0.267626211 | 0.445069003 | 2.86E-13 |
| CUPP | dbCAN_2:Hotpep | -0.550816385 | -0.639537781 | -0.462094989 | 2.86E-13 |
| CUPP | dbCAN_3 | -0.835396782 | -0.924118178 | -0.746675386 | 2.86E-13 |
| dbCAN_3:DIAMOND | dbCAN_3:HMMER | 0.356107255 | 0.267385859 | 0.44482865 | 2.86E-13 |
| dbCAN_4:DIAMOND | dbCAN_3:HMMER | 0.364695 | 0.275973604 | 0.453416396 | 2.86E-13 |
| CUPP | dbCAN_3:HMMER | -0.542468992 | -0.631190388 | -0.453747597 | 2.86E-13 |
| CUPP | dbCAN_3:DIAMOND | -0.898576247 | -0.987297643 | -0.809854851 | 2.86E-13 |
| CUPP | dbCAN_3:eCAMI | -0.680960908 | -0.769682304 | -0.592239513 | 2.86E-13 |
| CUPP | dbCAN_4 | -0.854719675 | -0.943441071 | -0.765998279 | 2.86E-13 |
| CUPP | dbCAN_4:HMMER | -0.595377139 | -0.684098534 | -0.506655743 | 2.86E-13 |
| CUPP | dbCAN_4:DIAMOND | -0.907163992 | -0.995885388 | -0.818442596 | 2.86E-13 |
| CUPP | dbCAN_4:dbCANsub | -0.857625673 | -0.946347069 | -0.768904277 | 2.86E-13 |
| dbCAN_4:dbCANsub | dbCAN_2:HMMER | 0.316250657 | 0.227529261 | 0.404972053 | 2.87E-13 |
| dbCAN_4:dbCANsub | dbCAN_3:HMMER | 0.31515668 | 0.226435285 | 0.403878076 | 2.87E-13 |
| dbCAN_4 | dbCAN_2:HMMER | 0.313344659 | 0.224623264 | 0.402066055 | 2.87E-13 |
| dbCAN_4 | dbCAN_3:HMMER | 0.312250683 | 0.223529287 | 0.400972079 | 2.89E-13 |
| dbCAN_4:DIAMOND | dbCAN_4:HMMER | 0.311786853 | 0.223065458 | 0.400508249 | 2.89E-13 |
| dbCAN_4:dbCANsub | dbCAN_2:Hotpep | 0.306809288 | 0.218087892 | 0.395530684 | 2.98E-13 |
| dbCAN_4 | dbCAN_2:Hotpep | 0.30390329 | 0.215181894 | 0.392624686 | 3.07E-13 |
| dbCAN_4:HMMER | dbCAN_3:DIAMOND | -0.303199108 | -0.391920504 | -0.214477713 | 3.11E-13 |
| dbCAN_3 | dbCAN_2:HMMER | 0.294021766 | 0.20530037 | 0.382743162 | 3.65E-13 |
| dbCAN_3:HMMER | dbCAN_3 | -0.292927789 | -0.381649185 | -0.204206393 | 3.72E-13 |
| dbCAN_2:DIAMOND | dbCAN_2:HMMER | 0.291157699 | 0.202436303 | 0.379879095 | 3.84E-13 |
| dbCAN_3:HMMER | dbCAN_2:DIAMOND | -0.290063723 | -0.378785119 | -0.201342327 | 3.88E-13 |
| dbCAN_4:HMMER | dbCAN_2:DIAMOND | -0.237155576 | -0.325876972 | -0.148434181 | 3.96E-13 |
| dbCAN_3 | dbCAN_2:Hotpep | 0.284580397 | 0.195859001 | 0.373301793 | 4.02E-13 |
| dbCAN_4:HMMER | dbCAN_3 | -0.240019643 | -0.328741039 | -0.151298247 | 4.04E-13 |
| dbCAN_2:HMMER | dbCAN_2 | -0.255659494 | -0.344380889 | -0.166938098 | 4.06E-13 |
| dbCAN_4:dbCANsub | dbCAN_4:HMMER | 0.262248534 | 0.173527138 | 0.35096993 | 4.07E-13 |
| dbCAN_2:Hotpep | dbCAN_2:DIAMOND | -0.28171633 | -0.370437726 | -0.192994934 | 4.08E-13 |
| dbCAN_4:HMMER | dbCAN_4 | -0.259342537 | -0.348063932 | -0.170621141 | 4.09E-13 |
| dbCAN_3:HMMER | dbCAN_2 | -0.254565517 | -0.343286913 | -0.165844121 | 4.12E-13 |
| dbCAN_2:Hotpep | dbCAN_2 | -0.246218125 | -0.33493952 | -0.157496729 | 4.16E-13 |
| dbCAN_4:DIAMOND | dbCAN_3:eCAMI | 0.226203084 | 0.137481688 | 0.31492448 | 4.19E-13 |
| dbCAN_3:eCAMI | dbCAN_3:DIAMOND | -0.217615339 | -0.306336735 | -0.128893943 | 5.25E-13 |
| dbCAN_4:HMMER | dbCAN_2 | -0.201657371 | -0.290378767 | -0.112935975 | 7.98E-12 |
| dbCAN_4:dbCANsub | dbCAN_3:eCAMI | 0.176664764 | 0.087943369 | 0.26538616 | 4.74E-09 |
| dbCAN_4 | dbCAN_3:eCAMI | 0.173758767 | 0.085037371 | 0.262480163 | 9.53E-09 |
| dbCAN_3:eCAMI | dbCAN_3 | -0.154435873 | -0.243157269 | -0.065714477 | 7.65E-07 |
| dbCAN_3:eCAMI | dbCAN_2:DIAMOND | -0.151571807 | -0.240293203 | -0.062850411 | 1.41E-06 |
| dbCAN_3:eCAMI | dbCAN_2:HMMER | 0.139585893 | 0.050864497 | 0.228307288 | 1.61E-05 |
| dbCAN_3:eCAMI | dbCAN_3:HMMER | 0.138491916 | 0.04977052 | 0.227213312 | 2.00E-05 |
| dbCAN_3:eCAMI | dbCAN_2:Hotpep | 0.130144524 | 0.041423128 | 0.218865919 | 9.62E-05 |
| dbCAN_3:eCAMI | dbCAN_2 | -0.116073601 | -0.204794997 | -0.027352205 | 0.001091898 |
| dbCAN_4:DIAMOND | dbCAN_2 | 0.110129483 | 0.021408087 | 0.198850879 | 0.002787351 |
| dbCAN_3:DIAMOND | dbCAN_2 | 0.101541738 | 0.012820342 | 0.190263133 | 0.009767679 |
| dbCAN_4:HMMER | dbCAN_3:eCAMI | -0.08558377 | -0.174305166 | 0.003137626 | 0.071109394 |
| dbCAN_4:DIAMOND | dbCAN_2:DIAMOND | 0.074631277 | -0.014090119 | 0.163352673 | 0.206808439 |
| dbCAN_4:DIAMOND | dbCAN_3 | 0.07176721 | -0.016954185 | 0.160488606 | 0.261577507 |
| dbCAN_3:DIAMOND | dbCAN_2:DIAMOND | 0.066043532 | -0.022677864 | 0.154764928 | 0.394491273 |
| dbCAN_3:DIAMOND | dbCAN_3 | 0.063179465 | -0.025541931 | 0.151900861 | 0.470128397 |
| dbCAN_4:dbCANsub | dbCAN_2 | 0.060591163 | -0.028130232 | 0.149312559 | 0.541270183 |
| dbCAN_4 | dbCAN_2 | 0.057685166 | -0.03103623 | 0.146406562 | 0.621535867 |
| dbCAN_4:HMMER | dbCAN_2:HMMER | 0.054002123 | -0.034719273 | 0.142723519 | 0.71879984 |
| dbCAN_4:HMMER | dbCAN_3:HMMER | 0.052908146 | -0.03581325 | 0.141629542 | 0.745834351 |
| dbCAN_4:DIAMOND | dbCAN_4 | 0.052444317 | -0.036277079 | 0.141165713 | 0.756962835 |
| dbCAN_4:dbCANsub | dbCAN_4:DIAMOND | -0.049538319 | -0.138259715 | 0.039183077 | 0.821423698 |
| dbCAN_4:HMMER | dbCAN_2:Hotpep | 0.044560754 | -0.044160642 | 0.13328215 | 0.90707641 |
| dbCAN_4 | dbCAN_3:DIAMOND | -0.043856572 | -0.132577968 | 0.044864824 | 0.916494296 |
| dbCAN_4:dbCANsub | dbCAN_3:DIAMOND | -0.040950574 | -0.12967197 | 0.047770822 | 0.948530096 |
| dbCAN_3 | dbCAN_2 | 0.038362272 | -0.050359124 | 0.127083668 | 0.968623106 |
| dbCAN_2:DIAMOND | dbCAN_2 | 0.035498206 | -0.05322319 | 0.124219602 | 0.983248708 |
| dbCAN_4:dbCANsub | dbCAN_2:DIAMOND | 0.025092958 | -0.063628438 | 0.113814354 | 0.999330902 |
| dbCAN_4:dbCANsub | dbCAN_3 | 0.022228891 | -0.066492505 | 0.110950287 | 0.999807179 |
| dbCAN_4 | dbCAN_2:DIAMOND | 0.02218696 | -0.066534436 | 0.110908356 | 0.99981095 |
| dbCAN_4 | dbCAN_3 | 0.019322893 | -0.069398502 | 0.108044289 | 0.999956617 |
| dbCAN_2:Hotpep | dbCAN_2:HMMER | 0.009441369 | -0.079280027 | 0.098162765 | 0.999999987 |
| dbCAN_4:DIAMOND | dbCAN_3:DIAMOND | 0.008587745 | -0.080133651 | 0.097309141 | 0.999999996 |
| dbCAN_3:HMMER | dbCAN_2:Hotpep | -0.008347392 | -0.097068788 | 0.080374003 | 0.999999997 |
| dbCAN_3:HMMER | dbCAN_2:HMMER | 0.001093977 | -0.087627419 | 0.089815372 | 1 |
| dbCAN_3 | dbCAN_2:DIAMOND | 0.002864067 | -0.085857329 | 0.091585462 | 1 |
| dbCAN_4:dbCANsub | dbCAN_4 | 0.002905998 | -0.085815398 | 0.091627394 | 1 |

**SI table 49: Output of Tukey HSD test for statistically significant differences between the mean accuracy of tools classifying CBM CAZyme domains (overleaf)**

| Classifier 1 | Classifier 2 | Mean Difference | Lower 95% Confidence Interval | Upper 95% Confidence Interval | Adjusted P-value |
|---|---|---|---|---|---|
| dbCAN_2:Hotpep | dbCAN_2 | -0.083272318 | -0.10368775 | -0.062856886 | 2.86E-13 |
| CUPP | dbCAN_2 | -0.080495683 | -0.100911115 | -0.060080251 | 2.86E-13 |
| dbCAN_2:Hotpep | dbCAN_2:DIAMOND | -0.089601238 | -0.11001667 | -0.069185806 | 2.86E-13 |
| CUPP | dbCAN_2:DIAMOND | -0.086824603 | -0.107240035 | -0.066409171 | 2.86E-13 |
| dbCAN_3 | dbCAN_2:Hotpep | 0.089423704 | 0.069008272 | 0.109839136 | 2.86E-13 |
| dbCAN_3:DIAMOND | dbCAN_2:Hotpep | 0.09908287 | 0.078667437 | 0.119498302 | 2.86E-13 |
| dbCAN_4 | dbCAN_2:Hotpep | 0.093976752 | 0.07356132 | 0.114392184 | 2.86E-13 |
| dbCAN_4:DIAMOND | dbCAN_2:Hotpep | 0.100661512 | 0.08024608 | 0.121076944 | 2.86E-13 |
| dbCAN_4:dbCANsub | dbCAN_2:Hotpep | 0.093815881 | 0.073400449 | 0.114231313 | 2.86E-13 |
| CUPP | dbCAN_3 | -0.086647069 | -0.107062501 | -0.066231637 | 2.86E-13 |
| CUPP | dbCAN_3:DIAMOND | -0.096306235 | -0.116721667 | -0.075890803 | 2.86E-13 |
| CUPP | dbCAN_4 | -0.091200117 | -0.111615549 | -0.070784685 | 2.86E-13 |
| CUPP | dbCAN_4:DIAMOND | -0.097884877 | -0.118300309 | -0.077469445 | 2.86E-13 |
| CUPP | dbCAN_4:dbCANsub | -0.091039246 | -0.111454678 | -0.070623814 | 2.86E-13 |
| dbCAN_3:eCAMI | dbCAN_3:DIAMOND | -0.054757111 | -0.075172543 | -0.034341678 | 4.04E-13 |
| CUPP | dbCAN_4:HMMER | -0.058529446 | -0.078944878 | -0.038114014 | 4.05E-13 |
| CUPP | dbCAN_3:HMMER | -0.052822769 | -0.073238201 | -0.032407337 | 4.05E-13 |
| dbCAN_2:Hotpep | dbCAN_2:HMMER | -0.055473307 | -0.075888739 | -0.035057875 | 4.07E-13 |
| dbCAN_3:HMMER | dbCAN_2:Hotpep | 0.055599404 | 0.035183972 | 0.076014836 | 4.07E-13 |
| CUPP | dbCAN_2:HMMER | -0.052696672 | -0.073112104 | -0.03228124 | 4.08E-13 |
| dbCAN_4:HMMER | dbCAN_2:Hotpep | 0.061306081 | 0.040890649 | 0.081721513 | 4.10E-13 |
| dbCAN_4:DIAMOND | dbCAN_3:eCAMI | 0.056335753 | 0.03592032 | 0.076751185 | 4.11E-13 |
| dbCAN_4 | dbCAN_3:eCAMI | 0.049650993 | 0.029235561 | 0.070066425 | 5.85E-13 |
| dbCAN_4:dbCANsub | dbCAN_3:eCAMI | 0.049490122 | 0.02907469 | 0.069905554 | 6.20E-13 |
| dbCAN_3:eCAMI | dbCAN_2:DIAMOND | -0.045275479 | -0.065690911 | -0.024860047 | 2.88E-11 |
| dbCAN_4:DIAMOND | dbCAN_2:HMMER | 0.045188205 | 0.024772773 | 0.065603637 | 3.19E-11 |
| dbCAN_3:eCAMI | dbCAN_3 | -0.045097945 | -0.065513377 | -0.024682513 | 3.53E-11 |
| dbCAN_4:DIAMOND | dbCAN_3:HMMER | 0.045062108 | 0.024646675 | 0.06547754 | 3.68E-11 |
| dbCAN_3:eCAMI | dbCAN_2:Hotpep | 0.044325759 | 0.023910327 | 0.064741191 | 8.50E-11 |
| dbCAN_3:DIAMOND | dbCAN_2:HMMER | 0.043609563 | 0.023194131 | 0.064024995 | 1.90E-10 |
| dbCAN_3:DIAMOND | dbCAN_3:HMMER | 0.043483465 | 0.023068033 | 0.063898898 | 2.19E-10 |
| CUPP | dbCAN_3:eCAMI | -0.041549124 | -0.061964556 | -0.021133692 | 1.82E-09 |
| dbCAN_4:DIAMOND | dbCAN_4:HMMER | 0.039355431 | 0.018939998 | 0.059770863 | 1.82E-08 |
| dbCAN_3:eCAMI | dbCAN_2 | -0.038946559 | -0.059361991 | -0.018531127 | 2.76E-08 |
| dbCAN_4 | dbCAN_2:HMMER | 0.038503445 | 0.018088013 | 0.058918877 | 4.32E-08 |
| dbCAN_4 | dbCAN_3:HMMER | 0.038377348 | 0.017961916 | 0.05879278 | 4.90E-08 |
| dbCAN_4:dbCANsub | dbCAN_2:HMMER | 0.038342574 | 0.017927142 | 0.058758006 | 5.07E-08 |
| dbCAN_4:dbCANsub | dbCAN_3:HMMER | 0.038216477 | 0.017801044 | 0.058631909 | 5.75E-08 |
| dbCAN_4:HMMER | dbCAN_3:DIAMOND | -0.037776789 | -0.058192221 | -0.017361356 | 8.90E-08 |
| dbCAN_2:DIAMOND | dbCAN_2:HMMER | 0.034127931 | 0.013712499 | 0.054543363 | 2.78E-06 |
| dbCAN_3:HMMER | dbCAN_2:DIAMOND | -0.034001834 | -0.054417266 | -0.013586402 | 3.12E-06 |
| dbCAN_3 | dbCAN_2:HMMER | 0.033950397 | 0.013534965 | 0.054365829 | 3.27E-06 |
| dbCAN_3:HMMER | dbCAN_3 | -0.0338243 | -0.054239732 | -0.013408868 | 3.66E-06 |
| dbCAN_4:HMMER | dbCAN_4 | -0.032670671 | -0.053086103 | -0.012255239 | 1.01E-05 |
| dbCAN_4:dbCANsub | dbCAN_4:HMMER | 0.0325098 | 0.012094368 | 0.052925232 | 1.16E-05 |
| dbCAN_4:HMMER | dbCAN_2:DIAMOND | -0.028295157 | -0.048710589 | -0.007879725 | 0.000344485 |
| dbCAN_4:HMMER | dbCAN_3 | -0.028117623 | -0.048533055 | -0.007702191 | 0.000393344 |
| dbCAN_2:HMMER | dbCAN_2 | -0.027799011 | -0.048214443 | -0.007383579 | 0.000497969 |
| dbCAN_3:HMMER | dbCAN_2 | -0.027672914 | -0.048088346 | -0.007257482 | 0.000546265 |
| dbCAN_4:HMMER | dbCAN_2 | -0.021966237 | -0.042381669 | -0.001550805 | 0.022000485 |
| dbCAN_4:DIAMOND | dbCAN_2 | 0.017389194 | -0.003026238 | 0.037804626 | 0.190691619 |
| dbCAN_4:HMMER | dbCAN_3:eCAMI | 0.016980322 | -0.00343511 | 0.037395754 | 0.221966762 |
| dbCAN_3:DIAMOND | dbCAN_2 | 0.015810552 | -0.00460488 | 0.036225984 | 0.32901743 |
| dbCAN_3:eCAMI | dbCAN_3:HMMER | -0.011273645 | -0.031689077 | 0.009141787 | 0.83240453 |
| dbCAN_4:DIAMOND | dbCAN_3 | 0.011237808 | -0.009177624 | 0.03165324 | 0.835471442 |
| dbCAN_3:eCAMI | dbCAN_2:HMMER | -0.011147548 | -0.03156298 | 0.009267884 | 0.843057321 |
| dbCAN_4:DIAMOND | dbCAN_2:DIAMOND | 0.011060274 | -0.009355158 | 0.031475706 | 0.850201614 |
| dbCAN_4 | dbCAN_2 | 0.010704434 | -0.009710998 | 0.031119866 | 0.877348231 |
| dbCAN_4:dbCANsub | dbCAN_2 | 0.010543563 | -0.009871869 | 0.030958995 | 0.888559138 |
| dbCAN_3:DIAMOND | dbCAN_3 | 0.009659166 | -0.010756267 | 0.030074598 | 0.938435296 |
| dbCAN_3:DIAMOND | dbCAN_2:DIAMOND | 0.009481632 | -0.0109338 | 0.029897064 | 0.946143894 |
| dbCAN_4:dbCANsub | dbCAN_4:DIAMOND | -0.006845631 | -0.027261063 | 0.013569801 | 0.996501657 |
| dbCAN_4:DIAMOND | dbCAN_4 | 0.00668476 | -0.013730672 | 0.027100192 | 0.997202163 |
| dbCAN_2:DIAMOND | dbCAN_2 | 0.00632892 | -0.014086512 | 0.026744352 | 0.998343026 |
| dbCAN_3 | dbCAN_2 | 0.006151386 | -0.014264046 | 0.026566818 | 0.998744687 |
| dbCAN_4:HMMER | dbCAN_2:HMMER | 0.005832774 | -0.014582658 | 0.026248206 | 0.999259369 |
| dbCAN_4:HMMER | dbCAN_3:HMMER | 0.005706677 | -0.014708755 | 0.026122109 | 0.999405587 |
| dbCAN_4:dbCANsub | dbCAN_3:DIAMOND | -0.005266989 | -0.025682421 | 0.015148443 | 0.999738447 |
| dbCAN_4 | dbCAN_3:DIAMOND | -0.005106118 | -0.02552155 | 0.015309315 | 0.999810664 |
| dbCAN_4 | dbCAN_3 | 0.004553048 | -0.015862384 | 0.02496848 | 0.99994396 |
| dbCAN_4:dbCANsub | dbCAN_3 | 0.004392177 | -0.016023255 | 0.024807609 | 0.999962012 |
| dbCAN_4 | dbCAN_2:DIAMOND | 0.004375514 | -0.016039918 | 0.024790946 | 0.999963547 |
| dbCAN_4:dbCANsub | dbCAN_2:DIAMOND | 0.004214643 | -0.016200789 | 0.024630075 | 0.999975765 |
| CUPP | dbCAN_2:Hotpep | 0.002776635 | -0.017638797 | 0.023192067 | 0.999999781 |
| dbCAN_4:DIAMOND | dbCAN_3:DIAMOND | 0.001578642 | -0.01883679 | 0.021994074 | 1 |
| dbCAN_3:HMMER | dbCAN_2:HMMER | 0.000126097 | -0.020289335 | 0.020541529 | 1 |
| dbCAN_3 | dbCAN_2:DIAMOND | -0.000177534 | -0.020592966 | 0.020237898 | 1 |
| dbCAN_4:dbCANsub | dbCAN_4 | -0.000160871 | -0.020576303 | 0.020254561 | 1 |

**SI table 50: Output of Tukey HSD test for statistically significant differences between the mean sensitivity of tools classifying CBM CAZyme domains (overleaf)**

| Classifier 1 | Classifier 2 | Mean Difference | Lower 95% Confidence Interval | Upper 95% Confidence Interval | Adjusted P-value |
|---|---|---|---|---|---|
| CUPP | dbCAN_2 | -0.724176654 | -0.827037753 | -0.621315555 | 2.86E-13 |
| dbCAN_3:DIAMOND | dbCAN_2:HMMER | 0.454542378 | 0.351681279 | 0.557403477 | 2.86E-13 |
| dbCAN_4 | dbCAN_2:HMMER | 0.386537061 | 0.283675962 | 0.48939816 | 2.86E-13 |
| dbCAN_4:DIAMOND | dbCAN_2:HMMER | 0.464386493 | 0.361525394 | 0.567247592 | 2.86E-13 |
| dbCAN_4:dbCANsub | dbCAN_2:HMMER | 0.406375351 | 0.303514252 | 0.50923645 | 2.86E-13 |
| CUPP | dbCAN_2:HMMER | -0.412947362 | -0.515808461 | -0.310086263 | 2.86E-13 |
| CUPP | dbCAN_2:DIAMOND | -0.785433758 | -0.888294857 | -0.682572659 | 2.86E-13 |
| CUPP | dbCAN_2:Hotpep | -0.706130496 | -0.808991595 | -0.603269397 | 2.86E-13 |
| CUPP | dbCAN_3 | -0.768006349 | -0.870867448 | -0.66514525 | 2.86E-13 |
| dbCAN_3:DIAMOND | dbCAN_3:HMMER | 0.454000576 | 0.351139477 | 0.556861675 | 2.86E-13 |
| dbCAN_4 | dbCAN_3:HMMER | 0.385995259 | 0.28313416 | 0.488856358 | 2.86E-13 |
| dbCAN_4:DIAMOND | dbCAN_3:HMMER | 0.463844691 | 0.360983592 | 0.56670579 | 2.86E-13 |
| dbCAN_4:dbCANsub | dbCAN_3:HMMER | 0.405833549 | 0.30297245 | 0.508694648 | 2.86E-13 |
| CUPP | dbCAN_3:HMMER | -0.413489164 | -0.516350263 | -0.310628065 | 2.86E-13 |
| dbCAN_4:HMMER | dbCAN_3:DIAMOND | -0.393449623 | -0.496310722 | -0.290588524 | 2.86E-13 |
| CUPP | dbCAN_3:DIAMOND | -0.86748974 | -0.970350839 | -0.764628641 | 2.86E-13 |
| CUPP | dbCAN_3:eCAMI | -0.734584266 | -0.837445365 | -0.631723167 | 2.86E-13 |
| CUPP | dbCAN_4 | -0.799484423 | -0.902345522 | -0.696623324 | 2.86E-13 |
| dbCAN_4:DIAMOND | dbCAN_4:HMMER | 0.403293738 | 0.300432639 | 0.506154837 | 2.86E-13 |
| CUPP | dbCAN_4:HMMER | -0.474040117 | -0.576901216 | -0.371179018 | 2.86E-13 |
| CUPP | dbCAN_4:DIAMOND | -0.877333854 | -0.980194953 | -0.774472755 | 2.86E-13 |
| CUPP | dbCAN_4:dbCANsub | -0.819322713 | -0.922183812 | -0.716461614 | 2.86E-13 |
| dbCAN_2:DIAMOND | dbCAN_2:HMMER | 0.372486396 | 0.269625297 | 0.475347495 | 2.86E-13 |
| dbCAN_3:HMMER | dbCAN_2:DIAMOND | -0.371944594 | -0.474805693 | -0.269083495 | 2.86E-13 |
| dbCAN_3 | dbCAN_2:HMMER | 0.355058987 | 0.252197888 | 0.457920086 | 2.99E-13 |
| dbCAN_3:HMMER | dbCAN_3 | -0.354517185 | -0.457378284 | -0.251656086 | 3.00E-13 |
| dbCAN_4:dbCANsub | dbCAN_4:HMMER | 0.345282596 | 0.242421497 | 0.448143695 | 3.39E-13 |
| dbCAN_3:HMMER | dbCAN_2:Hotpep | -0.292641332 | -0.395502431 | -0.189780233 | 4.06E-13 |
| dbCAN_4:HMMER | dbCAN_4 | -0.325444306 | -0.428305405 | -0.222583207 | 4.06E-13 |
| dbCAN_3:eCAMI | dbCAN_3:HMMER | 0.321095102 | 0.218234003 | 0.423956201 | 4.08E-13 |
| dbCAN_3:eCAMI | dbCAN_2:HMMER | 0.321636904 | 0.218775805 | 0.424498003 | 4.08E-13 |
| dbCAN_4:HMMER | dbCAN_3 | -0.293966232 | -0.396827331 | -0.191105133 | 4.09E-13 |
| dbCAN_2:Hotpep | dbCAN_2:HMMER | 0.293183134 | 0.190322035 | 0.396044233 | 4.10E-13 |
| dbCAN_4:HMMER | dbCAN_2:DIAMOND | -0.311393641 | -0.41425474 | -0.208532542 | 4.10E-13 |
| dbCAN_2:HMMER | dbCAN_2 | -0.311229292 | -0.414090391 | -0.208368193 | 4.11E-13 |
| dbCAN_3:HMMER | dbCAN_2 | -0.31068749 | -0.413548589 | -0.207826391 | 4.12E-13 |
| dbCAN_4:HMMER | dbCAN_3:eCAMI | -0.260544149 | -0.363405248 | -0.15768305 | 4.28E-13 |
| dbCAN_4:HMMER | dbCAN_2 | -0.250136537 | -0.352997636 | -0.147275438 | 5.86E-13 |
| dbCAN_4:HMMER | dbCAN_2:Hotpep | -0.232090379 | -0.334951478 | -0.12922928 | 1.17E-11 |
| dbCAN_4:DIAMOND | dbCAN_2:Hotpep | 0.171203359 | 0.06834226 | 0.274064458 | 3.18E-06 |
| dbCAN_3:DIAMOND | dbCAN_2:Hotpep | 0.161359244 | 0.058498145 | 0.264220343 | 1.75E-05 |
| dbCAN_4:DIAMOND | dbCAN_2 | 0.1531572 | 0.050296101 | 0.256018299 | 6.71E-05 |
| dbCAN_3:DIAMOND | dbCAN_2 | 0.143313086 | 0.040451987 | 0.246174185 | 0.000307987 |
| dbCAN_4:DIAMOND | dbCAN_3:eCAMI | 0.142749589 | 0.03988849 | 0.245610688 | 0.000335013 |
| dbCAN_3:eCAMI | dbCAN_3:DIAMOND | -0.132905475 | -0.235766574 | -0.030044376 | 0.001376613 |
| dbCAN_4:dbCANsub | dbCAN_2:Hotpep | 0.113192217 | 0.010331118 | 0.216053316 | 0.016587087 |
| dbCAN_4:DIAMOND | dbCAN_3 | 0.109327506 | 0.006466407 | 0.212188605 | 0.025500151 |
| dbCAN_3:DIAMOND | dbCAN_3 | 0.099483391 | -0.003377708 | 0.20234449 | 0.069387411 |
| dbCAN_4:dbCANsub | dbCAN_2 | 0.095146058 | -0.007715041 | 0.198007157 | 0.103043035 |
| dbCAN_4 | dbCAN_2:Hotpep | 0.093353927 | -0.009507172 | 0.196215026 | 0.120295741 |
| dbCAN_4:DIAMOND | dbCAN_2:DIAMOND | 0.091900096 | -0.010961003 | 0.194761195 | 0.135877661 |
| dbCAN_4:dbCANsub | dbCAN_3:eCAMI | 0.084738447 | -0.018122652 | 0.187599546 | 0.235240402 |
| dbCAN_3:DIAMOND | dbCAN_2:DIAMOND | 0.082055982 | -0.020805117 | 0.184917081 | 0.282456412 |
| dbCAN_2:Hotpep | dbCAN_2:DIAMOND | -0.079303262 | -0.182164361 | 0.023557837 | 0.336277618 |
| dbCAN_4:DIAMOND | dbCAN_4 | 0.077849431 | -0.025011668 | 0.18071053 | 0.366694727 |
| dbCAN_4 | dbCAN_2 | 0.075307769 | -0.02755333 | 0.178168868 | 0.422683976 |
| dbCAN_4 | dbCAN_3:DIAMOND | -0.068005317 | -0.170866416 | 0.034855782 | 0.594837006 |
| dbCAN_4 | dbCAN_3:eCAMI | 0.064900158 | -0.037960942 | 0.167761257 | 0.66757594 |
| dbCAN_3 | dbCAN_2:Hotpep | 0.061875853 | -0.040985246 | 0.164736952 | 0.73454708 |
| dbCAN_2:DIAMOND | dbCAN_2 | 0.061257104 | -0.041603995 | 0.164118203 | 0.747568023 |
| dbCAN_4:HMMER | dbCAN_2:HMMER | 0.061092755 | -0.041768344 | 0.163953854 | 0.750981523 |
| dbCAN_4:HMMER | dbCAN_3:HMMER | 0.060550953 | -0.042310146 | 0.163412052 | 0.762094863 |
| dbCAN_4:dbCANsub | dbCAN_4:DIAMOND | -0.058011142 | -0.160872241 | 0.044849957 | 0.811064215 |
| dbCAN_4:dbCANsub | dbCAN_3 | 0.051316364 | -0.051544735 | 0.154177463 | 0.911150535 |
| dbCAN_3:eCAMI | dbCAN_2:DIAMOND | -0.050849492 | -0.153710591 | 0.052011607 | 0.916456313 |
| dbCAN_4:dbCANsub | dbCAN_3:DIAMOND | -0.048167028 | -0.151028127 | 0.054694071 | 0.942831705 |
| dbCAN_3 | dbCAN_2 | 0.043829695 | -0.059031404 | 0.146690794 | 0.972039144 |
| dbCAN_4:dbCANsub | dbCAN_2:DIAMOND | 0.033888955 | -0.068972144 | 0.136750054 | 0.997034388 |
| dbCAN_3:eCAMI | dbCAN_3 | -0.033422083 | -0.136283182 | 0.069439016 | 0.997398972 |
| dbCAN_4 | dbCAN_3 | 0.031478074 | -0.071383025 | 0.134339173 | 0.998538761 |
| dbCAN_3:eCAMI | dbCAN_2:Hotpep | 0.02845377 | -0.074407329 | 0.131314869 | 0.999465214 |
| dbCAN_4:dbCANsub | dbCAN_4 | 0.01983829 | -0.08302281 | 0.122699389 | 0.999988537 |
| dbCAN_2:Hotpep | dbCAN_2 | -0.018046158 | -0.120907257 | 0.084814941 | 0.999996011 |
| dbCAN_3 | dbCAN_2:DIAMOND | -0.017427409 | -0.120288508 | 0.08543369 | 0.999997306 |
| dbCAN_4 | dbCAN_2:DIAMOND | 0.014050665 | -0.088810434 | 0.116911764 | 0.999999769 |
| dbCAN_3:eCAMI | dbCAN_2 | 0.010407611 | -0.092453488 | 0.11326871 | 0.999999993 |
| dbCAN_4:DIAMOND | dbCAN_3:DIAMOND | 0.009844114 | -0.093016985 | 0.112705213 | 0.999999996 |
| dbCAN_3:HMMER | dbCAN_2:HMMER | 0.000541802 | -0.102319297 | 0.103402901 | 1 |

# 7 Taxonomic performance of CAZy class classification across all CAZy classes

This section of the SI presents the data and figures for evaluation the performance of CAZy class classification across all CAZy classes.

### SI Table 51: Summary of CAZy class classification across all CAZy classes per taxonomic kingdom (overleaf)

| Statistical Parameter | Prediction Tool | Kingdom | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | Bacteria | | | | All | | | | Eukaryote | | | |
| | | Lower CI | Mean | SD | Upper CI | Lower CI | Mean | SD | Upper CI | Lower CI | Mean | SD | Upper CI |
| Specificity | dbCAN_2 | 0.993 | 0.995 | 0.016 | 0.997 | 0.991 | 0.993 | 0.026 | 0.996 | 0.987 | 0.992 | 0.033 | 0.996 |
| | dbCAN_2:HMMER | 0.994 | 0.996 | 0.014 | 0.998 | 0.991 | 0.994 | 0.026 | 0.996 | 0.987 | 0.992 | 0.034 | 0.997 |
| | dbCAN_2:DIAMOND | 0.991 | 0.993 | 0.019 | 0.996 | 0.989 | 0.992 | 0.027 | 0.995 | 0.986 | 0.991 | 0.033 | 0.995 |
| | dbCAN_2:Hotpep | 0.962 | 0.970 | 0.051 | 0.977 | 0.969 | 0.973 | 0.049 | 0.978 | 0.970 | 0.977 | 0.047 | 0.983 |
| | dbCAN_3 | 0.993 | 0.995 | 0.016 | 0.997 | 0.991 | 0.993 | 0.026 | 0.996 | 0.987 | 0.992 | 0.034 | 0.996 |
| | dbCAN_3:HMMER | 0.994 | 0.996 | 0.013 | 0.998 | 0.992 | 0.994 | 0.026 | 0.997 | 0.987 | 0.992 | 0.034 | 0.997 |
| | dbCAN_3:DIAMOND | 0.990 | 0.993 | 0.019 | 0.996 | 0.988 | 0.991 | 0.029 | 0.994 | 0.984 | 0.989 | 0.036 | 0.994 |
| | dbCAN_3:eCAMI | 0.974 | 0.980 | 0.039 | 0.985 | 0.979 | 0.983 | 0.039 | 0.987 | 0.981 | 0.986 | 0.038 | 0.991 |
| | dbCAN_4 | 0.994 | 0.996 | 0.014 | 0.998 | 0.991 | 0.994 | 0.027 | 0.996 | 0.987 | 0.992 | 0.035 | 0.996 |
| | dbCAN_4:HMMER | 0.994 | 0.996 | 0.013 | 0.998 | 0.992 | 0.994 | 0.026 | 0.997 | 0.987 | 0.992 | 0.034 | 0.997 |
| | dbCAN_4:DIAMOND | 0.992 | 0.994 | 0.017 | 0.997 | 0.989 | 0.992 | 0.029 | 0.994 | 0.984 | 0.989 | 0.036 | 0.994 |
| | dbCAN_4:dbCAN-sub | 0.993 | 0.995 | 0.014 | 0.997 | 0.990 | 0.993 | 0.027 | 0.996 | 0.986 | 0.991 | 0.035 | 0.996 |
| | CUPP | 0.995 | 0.997 | 0.013 | 0.999 | 0.993 | 0.995 | 0.025 | 0.998 | 0.989 | 0.994 | 0.032 | 0.998 |
| Sensitivity | dbCAN_2 | 0.859 | 0.883 | 0.171 | 0.907 | 0.857 | 0.873 | 0.173 | 0.890 | 0.841 | 0.864 | 0.175 | 0.888 |
| | dbCAN_2:HMMER | 0.789 | 0.819 | 0.222 | 0.850 | 0.784 | 0.807 | 0.246 | 0.831 | 0.760 | 0.796 | 0.267 | 0.832 |
| | dbCAN_2:DIAMOND | 0.864 | 0.889 | 0.182 | 0.914 | 0.853 | 0.873 | 0.204 | 0.893 | 0.828 | 0.858 | 0.223 | 0.889 |
| | dbCAN_2:Hotpep | 0.788 | 0.818 | 0.211 | 0.847 | 0.782 | 0.802 | 0.205 | 0.822 | 0.761 | 0.788 | 0.198 | 0.815 |
| | dbCAN_3 | 0.890 | 0.913 | 0.161 | 0.935 | 0.905 | 0.919 | 0.145 | 0.933 | 0.907 | 0.924 | 0.129 | 0.942 |
| | dbCAN_3:HMMER | 0.795 | 0.826 | 0.221 | 0.856 | 0.802 | 0.825 | 0.244 | 0.849 | 0.789 | 0.825 | 0.265 | 0.861 |
| | dbCAN_3:DIAMOND | 0.914 | 0.933 | 0.143 | 0.953 | 0.937 | 0.950 | 0.128 | 0.962 | 0.950 | 0.965 | 0.109 | 0.980 |
| | dbCAN_3:eCAMI | 0.815 | 0.844 | 0.207 | 0.872 | 0.790 | 0.811 | 0.215 | 0.831 | 0.749 | 0.779 | 0.218 | 0.809 |
| | dbCAN_4 | 0.933 | 0.948 | 0.108 | 0.963 | 0.923 | 0.935 | 0.129 | 0.948 | 0.904 | 0.923 | 0.146 | 0.943 |
| | dbCAN_4:HMMER | 0.827 | 0.855 | 0.195 | 0.882 | 0.821 | 0.843 | 0.228 | 0.865 | 0.797 | 0.832 | 0.256 | 0.867 |
| | dbCAN_4:DIAMOND | 0.911 | 0.932 | 0.146 | 0.952 | 0.947 | 0.958 | 0.111 | 0.969 | 0.977 | 0.983 | 0.049 | 0.990 |
| | dbCAN_4:dbCAN-sub | 0.942 | 0.956 | 0.101 | 0.970 | 0.925 | 0.937 | 0.126 | 0.949 | 0.899 | 0.919 | 0.143 | 0.938 |
| | CUPP | 0.664 | 0.717 | 0.382 | 0.770 | 0.678 | 0.714 | 0.374 | 0.751 | 0.662 | 0.712 | 0.368 | 0.762 |
| Precision | dbCAN_2 | 0.949 | 0.966 | 0.122 | 0.983 | 0.944 | 0.957 | 0.129 | 0.969 | 0.929 | 0.948 | 0.136 | 0.966 |
| | dbCAN_2:HMMER | 0.957 | 0.971 | 0.097 | 0.984 | 0.941 | 0.954 | 0.143 | 0.968 | 0.915 | 0.939 | 0.175 | 0.963 |
| | dbCAN_2:DIAMOND | 0.946 | 0.965 | 0.135 | 0.983 | 0.933 | 0.949 | 0.157 | 0.964 | 0.910 | 0.933 | 0.174 | 0.957 |
| | dbCAN_2:Hotpep | 0.821 | 0.855 | 0.242 | 0.889 | 0.832 | 0.856 | 0.244 | 0.880 | 0.823 | 0.857 | 0.246 | 0.890 |
| | dbCAN_3 | 0.953 | 0.969 | 0.116 | 0.985 | 0.955 | 0.965 | 0.110 | 0.976 | 0.948 | 0.962 | 0.104 | 0.976 |
| | dbCAN_3:HMMER | 0.960 | 0.972 | 0.093 | 0.985 | 0.944 | 0.957 | 0.135 | 0.970 | 0.920 | 0.943 | 0.165 | 0.965 |
| | dbCAN_3:DIAMOND | 0.956 | 0.971 | 0.111 | 0.987 | 0.946 | 0.958 | 0.125 | 0.971 | 0.928 | 0.946 | 0.136 | 0.965 |
| | dbCAN_3:eCAMI | 0.857 | 0.886 | 0.214 | 0.916 | 0.867 | 0.888 | 0.218 | 0.909 | 0.859 | 0.889 | 0.222 | 0.920 |
| | dbCAN_4 | 0.968 | 0.978 | 0.075 | 0.989 | 0.959 | 0.968 | 0.094 | 0.978 | 0.944 | 0.959 | 0.109 | 0.974 |
| | dbCAN_4:HMMER | 0.964 | 0.975 | 0.085 | 0.987 | 0.950 | 0.962 | 0.124 | 0.974 | 0.928 | 0.948 | 0.151 | 0.969 |
| | dbCAN_4:DIAMOND | 0.958 | 0.973 | 0.110 | 0.989 | 0.952 | 0.963 | 0.115 | 0.974 | 0.937 | 0.954 | 0.119 | 0.970 |
| | dbCAN_4:dbCAN-sub | 0.966 | 0.977 | 0.075 | 0.987 | 0.955 | 0.964 | 0.100 | 0.974 | 0.936 | 0.952 | 0.118 | 0.969 |
| | CUPP | 0.725 | 0.781 | 0.402 | 0.837 | 0.740 | 0.779 | 0.398 | 0.817 | 0.723 | 0.777 | 0.396 | 0.830 |
| F1-score | dbCAN_2 | 0.892 | 0.912 | 0.142 | 0.932 | 0.889 | 0.903 | 0.144 | 0.916 | 0.874 | 0.893 | 0.146 | 0.913 |
| | dbCAN_2:HMMER | 0.844 | 0.868 | 0.174 | 0.892 | 0.831 | 0.851 | 0.204 | 0.871 | 0.804 | 0.835 | 0.228 | 0.866 |
| | dbCAN_2:DIAMOND | 0.892 | 0.914 | 0.154 | 0.935 | 0.879 | 0.896 | 0.176 | 0.913 | 0.853 | 0.879 | 0.194 | 0.906 |
| | dbCAN_2:Hotpep | 0.785 | 0.814 | 0.209 | 0.843 | 0.786 | 0.806 | 0.206 | 0.826 | 0.770 | 0.798 | 0.203 | 0.825 |
| | dbCAN_3 | 0.911 | 0.930 | 0.137 | 0.949 | 0.920 | 0.932 | 0.120 | 0.944 | 0.920 | 0.934 | 0.103 | 0.948 |
| | dbCAN_3:HMMER | 0.849 | 0.873 | 0.173 | 0.897 | 0.843 | 0.863 | 0.200 | 0.882 | 0.823 | 0.853 | 0.224 | 0.884 |
| | dbCAN_3:DIAMOND | 0.927 | 0.943 | 0.121 | 0.960 | 0.935 | 0.946 | 0.117 | 0.957 | 0.933 | 0.948 | 0.114 | 0.964 |
| | dbCAN_3:eCAMI | 0.823 | 0.851 | 0.198 | 0.878 | 0.813 | 0.832 | 0.202 | 0.852 | 0.787 | 0.815 | 0.205 | 0.843 |
| | dbCAN_4 | 0.945 | 0.957 | 0.085 | 0.969 | 0.933 | 0.943 | 0.104 | 0.953 | 0.914 | 0.930 | 0.118 | 0.946 |
| | dbCAN_4:HMMER | 0.874 | 0.895 | 0.149 | 0.916 | 0.859 | 0.877 | 0.186 | 0.895 | 0.830 | 0.859 | 0.215 | 0.888 |
| | dbCAN_4:DIAMOND | 0.926 | 0.943 | 0.124 | 0.960 | 0.943 | 0.953 | 0.105 | 0.963 | 0.952 | 0.963 | 0.082 | 0.974 |
| | dbCAN_4:dbCAN-sub | 0.949 | 0.961 | 0.081 | 0.972 | 0.932 | 0.942 | 0.105 | 0.953 | 0.908 | 0.925 | 0.121 | 0.941 |
| | CUPP | 0.687 | 0.741 | 0.386 | 0.795 | 0.702 | 0.739 | 0.380 | 0.776 | 0.686 | 0.737 | 0.375 | 0.788 |
| Accuracy | dbCAN_2 | 0.969 | 0.975 | 0.044 | 0.981 | 0.970 | 0.974 | 0.043 | 0.978 | 0.967 | 0.973 | 0.043 | 0.979 |
| | dbCAN_2:HMMER | 0.958 | 0.965 | 0.048 | 0.971 | 0.961 | 0.965 | 0.041 | 0.969 | 0.961 | 0.966 | 0.033 | 0.970 |
| | dbCAN_2:DIAMOND | 0.970 | 0.977 | 0.046 | 0.983 | 0.972 | 0.976 | 0.045 | 0.981 | 0.970 | 0.976 | 0.045 | 0.982 |
| | dbCAN_2:Hotpep | 0.927 | 0.937 | 0.066 | 0.946 | 0.932 | 0.938 | 0.068 | 0.945 | 0.931 | 0.940 | 0.070 | 0.950 |
| | dbCAN_3 | 0.976 | 0.981 | 0.041 | 0.987 | 0.979 | 0.982 | 0.034 | 0.985 | 0.979 | 0.983 | 0.026 | 0.986 |
| | dbCAN_3:HMMER | 0.959 | 0.966 | 0.048 | 0.972 | 0.963 | 0.967 | 0.041 | 0.971 | 0.965 | 0.969 | 0.033 | 0.974 |
| | dbCAN_3:DIAMOND | 0.979 | 0.984 | 0.039 | 0.990 | 0.983 | 0.986 | 0.032 | 0.989 | 0.985 | 0.988 | 0.024 | 0.992 |
| | dbCAN_3:eCAMI | 0.950 | 0.958 | 0.054 | 0.965 | 0.950 | 0.955 | 0.056 | 0.961 | 0.945 | 0.953 | 0.059 | 0.961 |
| | dbCAN_4 | 0.980 | 0.986 | 0.039 | 0.991 | 0.981 | 0.984 | 0.033 | 0.987 | 0.979 | 0.982 | 0.025 | 0.986 |
| | dbCAN_4:HMMER | 0.962 | 0.968 | 0.047 | 0.975 | 0.965 | 0.969 | 0.040 | 0.973 | 0.965 | 0.969 | 0.033 | 0.974 |
| | dbCAN_4:DIAMOND | 0.978 | 0.984 | 0.040 | 0.989 | 0.984 | 0.987 | 0.032 | 0.990 | 0.987 | 0.990 | 0.023 | 0.993 |
| | dbCAN_4:dbCAN-sub | 0.981 | 0.986 | 0.036 | 0.991 | 0.980 | 0.983 | 0.032 | 0.986 | 0.977 | 0.980 | 0.026 | 0.984 |
| | CUPP | 0.941 | 0.952 | 0.074 | 0.962 | 0.948 | 0.954 | 0.062 | 0.960 | 0.950 | 0.956 | 0.047 | 0.962 |

# 8 Taxonomic performance of CAZy class classification per CAZy class

This section presents the same data as section 6, but presents the data grouped by CAZy class to facilitate evaluating the performance of CAZy class classification for CAZy classes of interest.

## 8.1 Specificity

### SI Table 52: Specificity of CAZy class classification per CAZy CAZy class, per taxonomic kingdom (overleaf)

Specificity was calculated per test set, the mean (and standard deviation) and 95% confidence interval (CI) was calculated by pooling all test sets.

| CAZy class | Prediction Tool | Kingdom | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | Bacteria | | | | All | | | | Eukaryote | | | |
| | | Lower CI | Mean | SD | Upper CI | Lower CI | Mean | SD | Upper CI | Lower CI | Mean | SD | Upper CI |
| GH | dbCAN_2 | 0.989 | 0.993 | 0.014 | 0.998 | 0.986 | 0.992 | 0.025 | 0.997 | 0.980 | 0.990 | 0.032 | 1.000 |
| | dbCAN_2:HMMER | 0.990 | 0.994 | 0.013 | 0.999 | 0.987 | 0.993 | 0.024 | 0.998 | 0.981 | 0.991 | 0.032 | 1.001 |
| | dbCAN_2:DIAMOND | 0.977 | 0.984 | 0.023 | 0.992 | 0.980 | 0.986 | 0.028 | 0.992 | 0.978 | 0.988 | 0.032 | 0.998 |
| | dbCAN_2:Hotpep | 0.966 | 0.975 | 0.028 | 0.984 | 0.977 | 0.983 | 0.029 | 0.990 | 0.983 | 0.991 | 0.027 | 1.000 |
| | dbCAN_3 | 0.989 | 0.993 | 0.014 | 0.998 | 0.987 | 0.992 | 0.023 | 0.997 | 0.982 | 0.992 | 0.029 | 1.001 |
| | dbCAN_3:HMMER | 0.991 | 0.995 | 0.012 | 0.999 | 0.989 | 0.994 | 0.022 | 0.999 | 0.983 | 0.992 | 0.029 | 1.002 |
| | dbCAN_3:DIAMOND | 0.977 | 0.985 | 0.024 | 0.992 | 0.978 | 0.984 | 0.029 | 0.991 | 0.973 | 0.984 | 0.034 | 0.995 |
| | dbCAN_3:eCAMI | 0.971 | 0.979 | 0.028 | 0.988 | 0.978 | 0.985 | 0.030 | 0.991 | 0.980 | 0.990 | 0.031 | 1.000 |
| | dbCAN_4 | 0.989 | 0.993 | 0.014 | 0.998 | 0.987 | 0.993 | 0.024 | 0.998 | 0.982 | 0.992 | 0.031 | 1.002 |
| | dbCAN_4:HMMER | 0.991 | 0.995 | 0.012 | 0.999 | 0.989 | 0.994 | 0.022 | 0.999 | 0.983 | 0.992 | 0.029 | 1.002 |
| | dbCAN_4:DIAMOND | 0.983 | 0.989 | 0.020 | 0.995 | 0.980 | 0.986 | 0.028 | 0.993 | 0.973 | 0.984 | 0.034 | 0.995 |
| | dbCAN_4:dbCAN-sub | 0.988 | 0.993 | 0.014 | 0.997 | 0.987 | 0.992 | 0.024 | 0.998 | 0.982 | 0.992 | 0.031 | 1.002 |
| | CUPP | 0.989 | 0.993 | 0.014 | 0.998 | 0.989 | 0.993 | 0.021 | 0.998 | 0.985 | 0.993 | 0.027 | 1.002 |
| GT | dbCAN_2 | 0.998 | 0.999 | 0.004 | 1.000 | 0.983 | 0.993 | 0.045 | 1.003 | 0.966 | 0.986 | 0.064 | 1.007 |
| | dbCAN_2:HMMER | 0.998 | 0.999 | 0.004 | 1.000 | 0.980 | 0.990 | 0.049 | 1.001 | 0.960 | 0.982 | 0.068 | 1.004 |
| | dbCAN_2:DIAMOND | 0.996 | 0.999 | 0.007 | 1.001 | 0.982 | 0.992 | 0.046 | 1.002 | 0.965 | 0.985 | 0.064 | 1.006 |
| | dbCAN_2:Hotpep | 0.996 | 0.998 | 0.006 | 1.000 | 0.983 | 0.992 | 0.042 | 1.002 | 0.968 | 0.987 | 0.059 | 1.006 |
| | dbCAN_3 | 0.998 | 0.999 | 0.004 | 1.000 | 0.981 | 0.991 | 0.047 | 1.002 | 0.962 | 0.984 | 0.066 | 1.005 |
| | dbCAN_3:HMMER | 0.998 | 0.999 | 0.004 | 1.000 | 0.980 | 0.990 | 0.049 | 1.001 | 0.960 | 0.982 | 0.068 | 1.004 |
| | dbCAN_3:DIAMOND | 0.996 | 0.998 | 0.007 | 1.001 | 0.978 | 0.989 | 0.049 | 1.000 | 0.959 | 0.980 | 0.068 | 1.002 |
| | dbCAN_3:eCAMI | 0.995 | 0.998 | 0.008 | 1.000 | 0.982 | 0.992 | 0.044 | 1.002 | 0.967 | 0.987 | 0.061 | 1.006 |
| | dbCAN_4 | 0.997 | 0.999 | 0.005 | 1.000 | 0.979 | 0.990 | 0.049 | 1.001 | 0.959 | 0.981 | 0.069 | 1.003 |
| | dbCAN_4:HMMER | 0.998 | 0.999 | 0.004 | 1.000 | 0.980 | 0.990 | 0.049 | 1.001 | 0.960 | 0.982 | 0.068 | 1.004 |
| | dbCAN_4:DIAMOND | 0.996 | 0.998 | 0.007 | 1.001 | 0.978 | 0.989 | 0.049 | 1.000 | 0.959 | 0.980 | 0.068 | 1.002 |
| | dbCAN_4:dbCAN-sub | 0.997 | 0.999 | 0.005 | 1.000 | 0.979 | 0.990 | 0.049 | 1.001 | 0.959 | 0.981 | 0.069 | 1.003 |
| | CUPP | 0.998 | 0.999 | 0.004 | 1.001 | 0.982 | 0.992 | 0.046 | 1.002 | 0.964 | 0.985 | 0.065 | 1.006 |
| PL | dbCAN_2 | 1.000 | 1.000 | 0.000 | 1.000 | 0.999 | 1.000 | 0.001 | 1.000 | 0.998 | 0.999 | 0.002 | 1.001 |
| | dbCAN_2:HMMER | 1.000 | 1.000 | 0.000 | 1.000 | 0.999 | 1.000 | 0.001 | 1.000 | 0.998 | 0.999 | 0.002 | 1.001 |
| | dbCAN_2:DIAMOND | 0.999 | 1.000 | 0.002 | 1.000 | 0.999 | 1.000 | 0.002 | 1.000 | 0.998 | 0.999 | 0.002 | 1.001 |
| | dbCAN_2:Hotpep | 0.998 | 0.999 | 0.003 | 1.000 | 0.999 | 0.999 | 0.003 | 1.000 | 0.998 | 0.999 | 0.002 | 1.001 |
| | dbCAN_3 | 0.999 | 1.000 | 0.002 | 1.000 | 0.999 | 0.999 | 0.003 | 1.000 | 0.997 | 0.999 | 0.003 | 1.000 |
| | dbCAN_3:HMMER | 0.999 | 1.000 | 0.002 | 1.000 | 0.999 | 0.999 | 0.003 | 1.000 | 0.997 | 0.999 | 0.003 | 1.000 |
| | dbCAN_3:DIAMOND | 0.998 | 0.999 | 0.003 | 1.000 | 0.998 | 0.999 | 0.004 | 1.000 | 0.995 | 0.998 | 0.005 | 1.001 |
| | dbCAN_3:eCAMI | 0.999 | 1.000 | 0.002 | 1.000 | 0.999 | 1.000 | 0.002 | 1.000 | 0.998 | 0.999 | 0.002 | 1.001 |
| | dbCAN_4 | 0.999 | 1.000 | 0.002 | 1.000 | 0.999 | 0.999 | 0.002 | 1.000 | 0.997 | 0.999 | 0.003 | 1.000 |
| | dbCAN_4:HMMER | 0.999 | 1.000 | 0.002 | 1.000 | 0.999 | 0.999 | 0.003 | 1.000 | 0.997 | 0.999 | 0.003 | 1.000 |
| | dbCAN_4:DIAMOND | 0.998 | 0.999 | 0.003 | 1.000 | 0.998 | 0.999 | 0.004 | 1.000 | 0.996 | 0.998 | 0.005 | 1.001 |
| | dbCAN_4:dbCAN-sub | 0.999 | 1.000 | 0.002 | 1.000 | 0.999 | 0.999 | 0.002 | 1.000 | 0.997 | 0.999 | 0.003 | 1.000 |
| | CUPP | 1.000 | 1.000 | 0.000 | 1.000 | 0.999 | 1.000 | 0.002 | 1.000 | 0.997 | 0.999 | 0.003 | 1.001 |
| CE | dbCAN_2 | 0.979 | 0.989 | 0.029 | 0.998 | 0.989 | 0.994 | 0.022 | 0.999 | 0.998 | 0.999 | 0.004 | 1.001 |
| | dbCAN_2:HMMER | 0.983 | 0.991 | 0.025 | 0.999 | 0.990 | 0.994 | 0.019 | 0.999 | 0.997 | 0.999 | 0.004 | 1.000 |
| | dbCAN_2:DIAMOND | 0.980 | 0.989 | 0.028 | 0.998 | 0.989 | 0.994 | 0.021 | 0.999 | 0.999 | 1.000 | 0.002 | 1.000 |
| | dbCAN_2:Hotpep | 0.973 | 0.983 | 0.030 | 0.992 | 0.985 | 0.990 | 0.023 | 0.996 | 0.997 | 0.999 | 0.004 | 1.000 |
| | dbCAN_3 | 0.979 | 0.988 | 0.030 | 0.998 | 0.989 | 0.994 | 0.022 | 0.999 | 0.998 | 0.999 | 0.004 | 1.001 |
| | dbCAN_3:HMMER | 0.984 | 0.992 | 0.024 | 1.000 | 0.991 | 0.995 | 0.018 | 0.999 | 0.998 | 0.999 | 0.004 | 1.000 |
| | dbCAN_3:DIAMOND | 0.980 | 0.989 | 0.028 | 0.999 | 0.987 | 0.993 | 0.023 | 0.998 | 0.991 | 0.996 | 0.016 | 1.001 |
| | dbCAN_3:eCAMI | 0.977 | 0.986 | 0.028 | 0.995 | 0.988 | 0.993 | 0.021 | 0.997 | 0.999 | 1.000 | 0.002 | 1.000 |
| | dbCAN_4 | 0.984 | 0.992 | 0.024 | 1.000 | 0.991 | 0.995 | 0.018 | 0.999 | 0.998 | 0.999 | 0.004 | 1.000 |
| | dbCAN_4:HMMER | 0.984 | 0.992 | 0.024 | 1.000 | 0.991 | 0.995 | 0.018 | 0.999 | 0.998 | 0.999 | 0.004 | 1.000 |
| | dbCAN_4:DIAMOND | 0.982 | 0.990 | 0.028 | 0.999 | 0.988 | 0.993 | 0.023 | 0.998 | 0.991 | 0.996 | 0.016 | 1.001 |
| | dbCAN_4:dbCAN-sub | 0.984 | 0.992 | 0.024 | 1.000 | 0.991 | 0.995 | 0.018 | 0.999 | 0.997 | 0.999 | 0.005 | 1.001 |
| | CUPP | 0.984 | 0.992 | 0.024 | 1.000 | 0.991 | 0.995 | 0.018 | 1.000 | 0.998 | 0.999 | 0.004 | 1.001 |
| AA | dbCAN_2 | 1.000 | 1.000 | 0.000 | 1.000 | 0.987 | 0.992 | 0.018 | 0.997 | 0.983 | 0.990 | 0.020 | 0.996 |
| | dbCAN_2:HMMER | 1.000 | 1.000 | 0.000 | 1.000 | 0.986 | 0.991 | 0.019 | 0.997 | 0.982 | 0.989 | 0.021 | 0.996 |
| | dbCAN_2:DIAMOND | 1.000 | 1.000 | 0.000 | 1.000 | 0.987 | 0.992 | 0.018 | 0.997 | 0.983 | 0.990 | 0.020 | 0.996 |
| | dbCAN_2:Hotpep | 1.000 | 1.000 | 0.000 | 1.000 | 0.987 | 0.992 | 0.018 | 0.997 | 0.983 | 0.990 | 0.020 | 0.996 |
| | dbCAN_3 | 1.000 | 1.000 | 0.000 | 1.000 | 0.985 | 0.991 | 0.020 | 0.997 | 0.981 | 0.988 | 0.022 | 0.996 |
| | dbCAN_3:HMMER | 1.000 | 1.000 | 0.000 | 1.000 | 0.985 | 0.991 | 0.020 | 0.997 | 0.980 | 0.988 | 0.022 | 0.995 |
| | dbCAN_3:DIAMOND | 1.000 | 1.000 | 0.000 | 1.000 | 0.984 | 0.990 | 0.022 | 0.996 | 0.979 | 0.987 | 0.024 | 0.995 |
| | dbCAN_3:eCAMI | 1.000 | 1.000 | 0.000 | 1.000 | 0.987 | 0.992 | 0.017 | 0.997 | 0.984 | 0.990 | 0.018 | 0.996 |
| | dbCAN_4 | 1.000 | 1.000 | 0.000 | 1.000 | 0.985 | 0.991 | 0.020 | 0.997 | 0.980 | 0.988 | 0.022 | 0.995 |
| | dbCAN_4:HMMER | 1.000 | 1.000 | 0.000 | 1.000 | 0.985 | 0.991 | 0.020 | 0.997 | 0.980 | 0.988 | 0.022 | 0.995 |
| | dbCAN_4:DIAMOND | 1.000 | 1.000 | 0.000 | 1.000 | 0.984 | 0.990 | 0.022 | 0.997 | 0.979 | 0.987 | 0.024 | 0.995 |
| | dbCAN_4:dbCAN-sub | 1.000 | 1.000 | 0.000 | 1.000 | 0.985 | 0.990 | 0.020 | 0.996 | 0.980 | 0.988 | 0.022 | 0.995 |
| | CUPP | 1.000 | 1.000 | 0.000 | 1.000 | 0.987 | 0.992 | 0.017 | 0.997 | 0.983 | 0.990 | 0.019 | 0.996 |
| CBM | dbCAN_2 | 0.992 | 0.995 | 0.009 | 0.998 | 0.990 | 0.992 | 0.012 | 0.995 | 0.985 | 0.990 | 0.014 | 0.994 |
| | dbCAN_2:HMMER | 0.992 | 0.995 | 0.011 | 0.998 | 0.994 | 0.996 | 0.009 | 0.998 | 0.995 | 0.997 | 0.007 | 0.999 |
| | dbCAN_2:DIAMOND | 0.990 | 0.994 | 0.015 | 0.999 | 0.988 | 0.991 | 0.015 | 0.995 | 0.984 | 0.989 | 0.015 | 0.993 |
| | dbCAN_2:Hotpep | 0.873 | 0.892 | 0.058 | 0.910 | 0.890 | 0.901 | 0.051 | 0.913 | 0.898 | 0.911 | 0.042 | 0.924 |
| | dbCAN_3 | 0.991 | 0.995 | 0.011 | 0.998 | 0.992 | 0.994 | 0.011 | 0.997 | 0.991 | 0.994 | 0.011 | 0.997 |
| | dbCAN_3:HMMER | 0.992 | 0.995 | 0.011 | 0.998 | 0.994 | 0.996 | 0.009 | 0.998 | 0.995 | 0.997 | 0.007 | 0.999 |
| | dbCAN_3:DIAMOND | 0.988 | 0.993 | 0.016 | 0.998 | 0.990 | 0.993 | 0.014 | 0.996 | 0.989 | 0.993 | 0.012 | 0.997 |
| | dbCAN_3:eCAMI | 0.916 | 0.935 | 0.059 | 0.954 | 0.935 | 0.947 | 0.052 | 0.959 | 0.946 | 0.959 | 0.042 | 0.973 |
| | dbCAN_4 | 0.992 | 0.995 | 0.011 | 0.999 | 0.993 | 0.995 | 0.010 | 0.997 | 0.992 | 0.995 | 0.010 | 0.998 |
| | dbCAN_4:HMMER | 0.992 | 0.995 | 0.011 | 0.998 | 0.994 | 0.996 | 0.010 | 0.998 | 0.994 | 0.997 | 0.009 | 0.999 |
| | dbCAN_4:DIAMOND | 0.989 | 0.994 | 0.015 | 0.999 | 0.991 | 0.994 | 0.015 | 0.997 | 0.990 | 0.994 | 0.014 | 0.998 |
| | dbCAN_4:dbCAN-sub | 0.990 | 0.994 | 0.011 | 0.998 | 0.990 | 0.993 | 0.013 | 0.996 | 0.987 | 0.991 | 0.014 | 0.996 |
| | CUPP | 1.000 | 1.000 | 0.000 | 1.000 | 1.000 | 1.000 | 0.000 | 1.000 | 1.000 | 1.000 | 0.000 | 1.000 |

## SI figure 24: The mean specificity and 95% confidence interval of the CAZy class classification across all CAZy classes (overleaf)

Specificity was calculated for each test set across all CAZy class. The mean specificity was then calculated across all test sets and 95% confidence interval (CI) was across all CAZy class, and is plotted in SI figure 24.

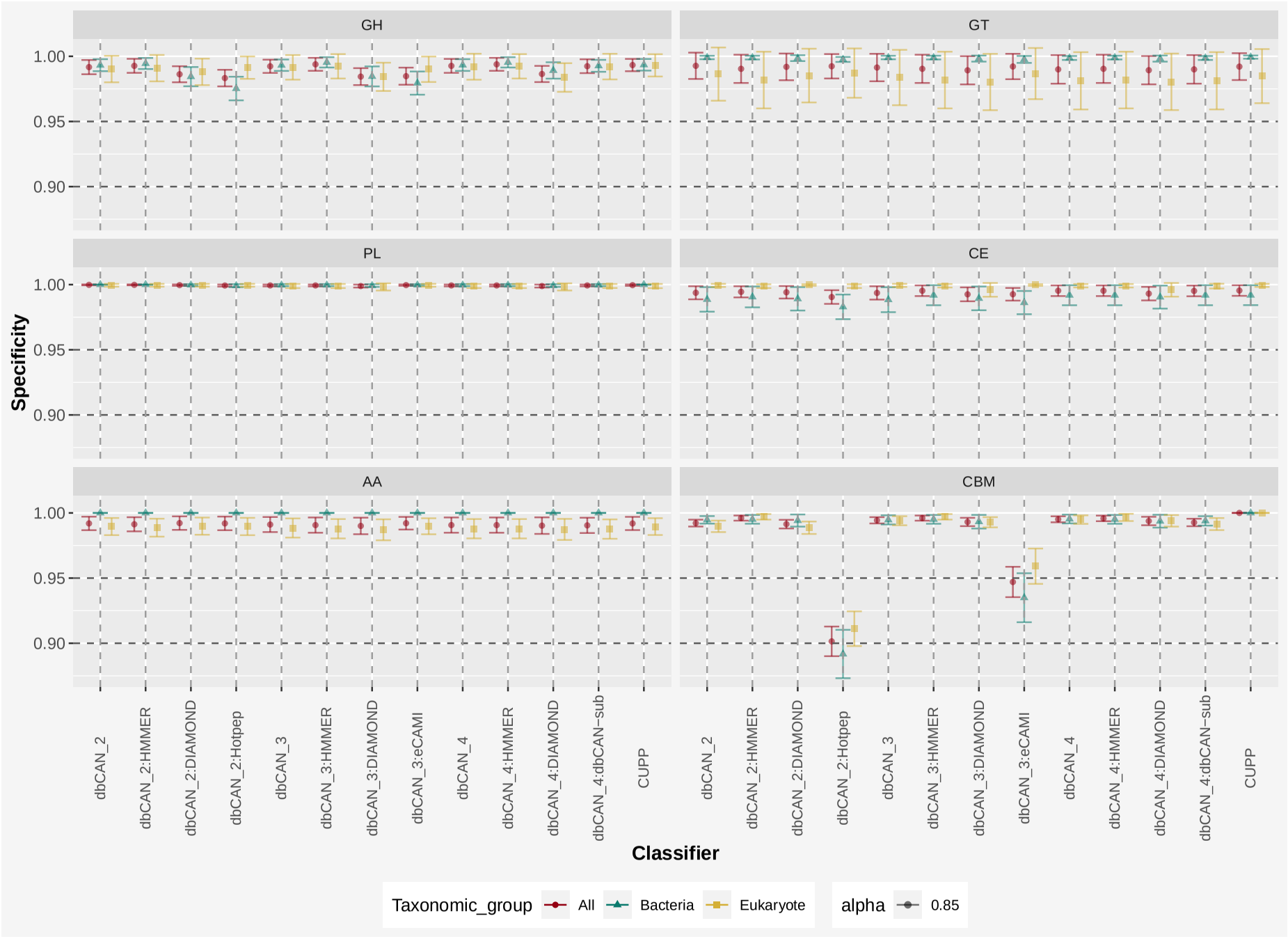



Figure 24: The mean specificity and 95% confidence interval (CI) of binary CAZy class classification across all CAZy classes and across all test sets. Where a tool is incorporated into a CAZyme classifier, the tool name is prefixed with the respective CAZyme classifier. Per classifier (from left to right) all test sets are pooled together (shaded red); all bacterial test sets are pooled (shaded green); and all test sets from eukaryotic genomes are pooled together (shaded yellow).

**SI figure 25: The specificity per test set of the CAZy class classification across all CAZy classes (overleaf)**

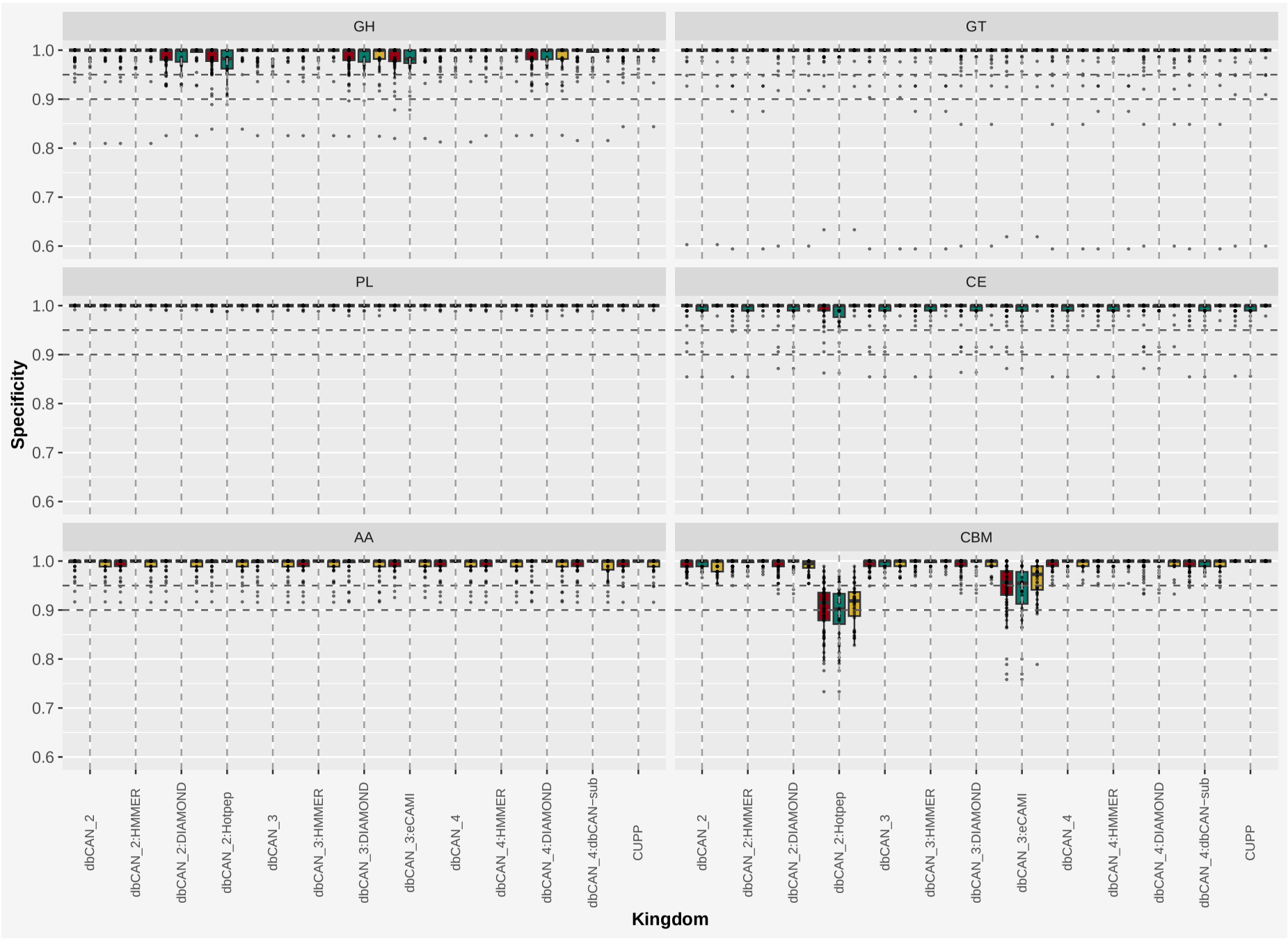



Figure 25: One-dimensional scatter plot overlaying a box and whisker plot, where each point represents the specificity of the binary CAZy class classification across all CAZy classes and per test set. Where a tool is incorporated into a CAZyme classifier, the tool name is prefixed with the respective CAZyme classifier. Per classifier (from left to right) all test sets are pooled together (shaded red); all bacterial test sets are pooled (shaded green); and all test sets from eukaryotic genomes are pooled together (shaded yellow).

## 8.2 Sensitivity

### SI Table 53: Sensitivity of CAZy class classification per CAZy CAZy class, per taxonomic kingdom (overleaf)

Sensitivity was calculated per test set, the mean (and standard deviation) and 95% confidence interval (CI) was calculated by pooling all test sets.

| CAZy class | Prediction Tool | Kingdom | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | Bacteria | | | | All | | | | Eukaryote | | | |
| | | Lower CI | Mean | SD | Upper CI | Lower CI | Mean | SD | Upper CI | Lower CI | Mean | SD | Upper CI |
| GH | dbCAN_2 | 0.904 | 0.936 | 0.102 | 0.969 | 0.915 | 0.935 | 0.091 | 0.956 | 0.909 | 0.934 | 0.079 | 0.960 |
| | dbCAN_2:HMMER | 0.877 | 0.912 | 0.110 | 0.947 | 0.890 | 0.908 | 0.082 | 0.926 | 0.891 | 0.904 | 0.041 | 0.917 |
| | dbCAN_2:DIAMOND | 0.897 | 0.934 | 0.116 | 0.971 | 0.914 | 0.938 | 0.106 | 0.962 | 0.910 | 0.942 | 0.097 | 0.973 |
| | dbCAN_2:Hotpep | 0.865 | 0.898 | 0.101 | 0.930 | 0.838 | 0.864 | 0.118 | 0.890 | 0.790 | 0.831 | 0.126 | 0.871 |
| | dbCAN_3 | 0.928 | 0.960 | 0.101 | 0.992 | 0.940 | 0.957 | 0.077 | 0.974 | 0.940 | 0.954 | 0.043 | 0.968 |
| | dbCAN_3:HMMER | 0.881 | 0.917 | 0.112 | 0.952 | 0.901 | 0.920 | 0.083 | 0.938 | 0.911 | 0.923 | 0.038 | 0.935 |
| | dbCAN_3:DIAMOND | 0.936 | 0.965 | 0.090 | 0.994 | 0.961 | 0.976 | 0.067 | 0.991 | 0.978 | 0.987 | 0.029 | 0.996 |
| | dbCAN_3:eCAMI | 0.880 | 0.914 | 0.106 | 0.948 | 0.852 | 0.876 | 0.110 | 0.901 | 0.806 | 0.839 | 0.101 | 0.871 |
| | dbCAN_4 | 0.919 | 0.953 | 0.109 | 0.988 | 0.932 | 0.950 | 0.080 | 0.968 | 0.936 | 0.947 | 0.034 | 0.958 |
| | dbCAN_4:HMMER | 0.882 | 0.917 | 0.112 | 0.953 | 0.902 | 0.920 | 0.083 | 0.939 | 0.911 | 0.923 | 0.038 | 0.935 |
| | dbCAN_4:DIAMOND | 0.925 | 0.959 | 0.105 | 0.992 | 0.957 | 0.974 | 0.076 | 0.991 | 0.982 | 0.989 | 0.019 | 0.995 |
| | dbCAN_4:dbCAN-sub | 0.921 | 0.955 | 0.106 | 0.989 | 0.930 | 0.947 | 0.079 | 0.965 | 0.928 | 0.940 | 0.036 | 0.951 |
| | CUPP | 0.892 | 0.917 | 0.079 | 0.942 | 0.893 | 0.908 | 0.067 | 0.923 | 0.882 | 0.899 | 0.053 | 0.916 |
| GT | dbCAN_2 | 0.837 | 0.882 | 0.139 | 0.927 | 0.854 | 0.884 | 0.138 | 0.915 | 0.843 | 0.887 | 0.138 | 0.931 |
| | dbCAN_2:HMMER | 0.807 | 0.849 | 0.131 | 0.891 | 0.838 | 0.863 | 0.113 | 0.888 | 0.847 | 0.876 | 0.090 | 0.905 |
| | dbCAN_2:DIAMOND | 0.873 | 0.924 | 0.158 | 0.975 | 0.892 | 0.926 | 0.151 | 0.959 | 0.881 | 0.927 | 0.145 | 0.973 |
| | dbCAN_2:Hotpep | 0.619 | 0.678 | 0.187 | 0.738 | 0.685 | 0.725 | 0.181 | 0.766 | 0.720 | 0.772 | 0.163 | 0.825 |
| | dbCAN_3 | 0.892 | 0.932 | 0.125 | 0.972 | 0.920 | 0.942 | 0.097 | 0.964 | 0.933 | 0.952 | 0.058 | 0.970 |
| | dbCAN_3:HMMER | 0.807 | 0.848 | 0.130 | 0.890 | 0.841 | 0.865 | 0.111 | 0.890 | 0.855 | 0.883 | 0.087 | 0.910 |
| | dbCAN_3:DIAMOND | 0.927 | 0.966 | 0.122 | 1.005 | 0.957 | 0.977 | 0.090 | 0.997 | 0.978 | 0.989 | 0.034 | 1.000 |
| | dbCAN_3:eCAMI | 0.803 | 0.855 | 0.162 | 0.907 | 0.816 | 0.850 | 0.152 | 0.884 | 0.799 | 0.845 | 0.144 | 0.891 |
| | dbCAN_4 | 0.905 | 0.945 | 0.126 | 0.985 | 0.937 | 0.958 | 0.092 | 0.978 | 0.961 | 0.971 | 0.032 | 0.981 |
| | dbCAN_4:HMMER | 0.808 | 0.849 | 0.130 | 0.891 | 0.841 | 0.866 | 0.111 | 0.890 | 0.854 | 0.882 | 0.087 | 0.910 |
| | dbCAN_4:DIAMOND | 0.914 | 0.955 | 0.126 | 0.995 | 0.955 | 0.975 | 0.091 | 0.995 | 0.992 | 0.996 | 0.012 | 0.999 |
| | dbCAN_4:dbCAN-sub | 0.909 | 0.945 | 0.112 | 0.980 | 0.935 | 0.954 | 0.083 | 0.972 | 0.951 | 0.963 | 0.038 | 0.975 |
| | CUPP | 0.839 | 0.875 | 0.114 | 0.912 | 0.829 | 0.854 | 0.111 | 0.878 | 0.798 | 0.832 | 0.104 | 0.865 |
| PL | dbCAN_2 | 0.831 | 0.910 | 0.210 | 0.988 | 0.809 | 0.880 | 0.242 | 0.951 | 0.678 | 0.826 | 0.289 | 0.975 |
| | dbCAN_2:HMMER | 0.860 | 0.921 | 0.164 | 0.982 | 0.835 | 0.897 | 0.212 | 0.960 | 0.712 | 0.856 | 0.279 | 1.000 |
| | dbCAN_2:DIAMOND | 0.853 | 0.927 | 0.196 | 1.000 | 0.790 | 0.869 | 0.269 | 0.948 | 0.589 | 0.768 | 0.348 | 0.946 |
| | dbCAN_2:Hotpep | 0.801 | 0.874 | 0.196 | 0.947 | 0.765 | 0.841 | 0.258 | 0.917 | 0.607 | 0.782 | 0.341 | 0.958 |
| | dbCAN_3 | 0.947 | 0.981 | 0.092 | 1.016 | 0.967 | 0.988 | 0.073 | 1.010 | 1.000 | 1.000 | 0.000 | 1.000 |
| | dbCAN_3:HMMER | 0.914 | 0.959 | 0.121 | 1.004 | 0.945 | 0.974 | 0.098 | 1.003 | 1.000 | 1.000 | 0.000 | 1.000 |
| | dbCAN_3:DIAMOND | 0.994 | 0.998 | 0.011 | 1.002 | 0.967 | 0.988 | 0.073 | 1.010 | 0.908 | 0.971 | 0.121 | 1.033 |
| | dbCAN_3:eCAMI | 0.728 | 0.818 | 0.247 | 0.909 | 0.715 | 0.796 | 0.278 | 0.877 | 0.586 | 0.756 | 0.331 | 0.926 |
| | dbCAN_4 | 0.914 | 0.959 | 0.121 | 1.004 | 0.945 | 0.974 | 0.098 | 1.003 | 1.000 | 1.000 | 0.000 | 1.000 |
| | dbCAN_4:HMMER | 0.914 | 0.959 | 0.121 | 1.004 | 0.945 | 0.974 | 0.098 | 1.003 | 1.000 | 1.000 | 0.000 | 1.000 |
| | dbCAN_4:DIAMOND | 0.994 | 0.998 | 0.011 | 1.002 | 0.996 | 0.999 | 0.009 | 1.001 | 1.000 | 1.000 | 0.000 | 1.000 |
| | dbCAN_4:dbCAN-sub | 0.916 | 0.961 | 0.121 | 1.006 | 0.946 | 0.975 | 0.098 | 1.004 | 1.000 | 1.000 | 0.000 | 1.000 |
| | CUPP | 0.754 | 0.848 | 0.252 | 0.942 | 0.775 | 0.851 | 0.259 | 0.927 | 0.712 | 0.856 | 0.279 | 1.000 |
| CE | dbCAN_2 | 0.892 | 0.930 | 0.118 | 0.967 | 0.889 | 0.921 | 0.143 | 0.954 | 0.855 | 0.912 | 0.168 | 0.969 |
| | dbCAN_2:HMMER | 0.832 | 0.880 | 0.150 | 0.928 | 0.890 | 0.921 | 0.136 | 0.952 | 0.931 | 0.966 | 0.103 | 1.001 |
| | dbCAN_2:DIAMOND | 0.826 | 0.880 | 0.169 | 0.934 | 0.792 | 0.848 | 0.247 | 0.904 | 0.708 | 0.813 | 0.311 | 0.918 |
| | dbCAN_2:Hotpep | 0.883 | 0.923 | 0.125 | 0.963 | 0.804 | 0.851 | 0.206 | 0.898 | 0.687 | 0.771 | 0.248 | 0.855 |
| | dbCAN_3 | 0.871 | 0.921 | 0.157 | 0.971 | 0.893 | 0.928 | 0.154 | 0.963 | 0.885 | 0.936 | 0.152 | 0.988 |
| | dbCAN_3:HMMER | 0.832 | 0.880 | 0.150 | 0.928 | 0.892 | 0.923 | 0.133 | 0.953 | 0.939 | 0.971 | 0.093 | 1.002 |
| | dbCAN_3:DIAMOND | 0.891 | 0.934 | 0.135 | 0.977 | 0.891 | 0.930 | 0.170 | 0.969 | 0.857 | 0.926 | 0.205 | 0.995 |
| | dbCAN_3:eCAMI | 0.803 | 0.864 | 0.191 | 0.925 | 0.751 | 0.807 | 0.245 | 0.863 | 0.649 | 0.744 | 0.283 | 0.840 |
| | dbCAN_4 | 0.940 | 0.967 | 0.084 | 0.994 | 0.963 | 0.978 | 0.067 | 0.994 | 0.978 | 0.991 | 0.039 | 1.004 |
| | dbCAN_4:HMMER | 0.884 | 0.924 | 0.127 | 0.965 | 0.930 | 0.953 | 0.101 | 0.976 | 0.969 | 0.984 | 0.046 | 1.000 |
| | dbCAN_4:DIAMOND | 0.926 | 0.962 | 0.110 | 0.997 | 0.956 | 0.975 | 0.085 | 0.995 | 0.978 | 0.991 | 0.039 | 1.004 |
| | dbCAN_4:dbCAN-sub | 0.936 | 0.963 | 0.085 | 0.991 | 0.960 | 0.975 | 0.068 | 0.991 | 0.976 | 0.989 | 0.038 | 1.002 |
| | CUPP | 0.856 | 0.898 | 0.132 | 0.941 | 0.881 | 0.911 | 0.133 | 0.942 | 0.880 | 0.926 | 0.135 | 0.972 |
| AA | dbCAN_2 | 1.000 | 1.000 | 0.000 | 1.000 | 0.886 | 0.919 | 0.115 | 0.952 | 0.854 | 0.895 | 0.121 | 0.935 |
| | dbCAN_2:HMMER | 1.000 | 1.000 | 0.000 | 1.000 | 0.911 | 0.937 | 0.092 | 0.964 | 0.887 | 0.919 | 0.097 | 0.951 |
| | dbCAN_2:DIAMOND | 1.000 | 1.000 | 0.000 | 1.000 | 0.817 | 0.869 | 0.179 | 0.921 | 0.768 | 0.830 | 0.187 | 0.893 |
| | dbCAN_2:Hotpep | 0.707 | 0.909 | 0.302 | 1.112 | 0.817 | 0.874 | 0.194 | 0.930 | 0.812 | 0.863 | 0.153 | 0.914 |
| | dbCAN_3 | 1.000 | 1.000 | 0.000 | 1.000 | 0.970 | 0.984 | 0.045 | 0.997 | 0.962 | 0.979 | 0.050 | 0.995 |
| | dbCAN_3:HMMER | 1.000 | 1.000 | 0.000 | 1.000 | 0.977 | 0.988 | 0.039 | 0.999 | 0.970 | 0.984 | 0.044 | 0.999 |
| | dbCAN_3:DIAMOND | 1.000 | 1.000 | 0.000 | 1.000 | 0.978 | 0.988 | 0.036 | 0.999 | 0.971 | 0.985 | 0.040 | 0.998 |
| | dbCAN_3:eCAMI | 0.707 | 0.909 | 0.302 | 1.112 | 0.706 | 0.782 | 0.261 | 0.858 | 0.664 | 0.744 | 0.239 | 0.824 |
| | dbCAN_4 | 1.000 | 1.000 | 0.000 | 1.000 | 0.986 | 0.994 | 0.026 | 1.001 | 0.982 | 0.992 | 0.029 | 1.002 |
| | dbCAN_4:HMMER | 1.000 | 1.000 | 0.000 | 1.000 | 0.977 | 0.988 | 0.039 | 0.999 | 0.970 | 0.984 | 0.044 | 0.999 |
| | dbCAN_4:DIAMOND | 0.853 | 0.955 | 0.151 | 1.056 | 0.942 | 0.970 | 0.095 | 0.998 | 0.950 | 0.975 | 0.074 | 0.999 |
| | dbCAN_4:dbCAN-sub | 1.000 | 1.000 | 0.000 | 1.000 | 0.980 | 0.989 | 0.033 | 0.999 | 0.974 | 0.986 | 0.037 | 0.998 |
| | CUPP | 1.000 | 1.000 | 0.000 | 1.000 | 0.867 | 0.905 | 0.130 | 0.942 | 0.831 | 0.876 | 0.136 | 0.922 |
| CBM | dbCAN_2 | 0.668 | 0.732 | 0.200 | 0.796 | 0.681 | 0.724 | 0.192 | 0.767 | 0.657 | 0.717 | 0.187 | 0.776 |
| | dbCAN_2:HMMER | 0.438 | 0.511 | 0.228 | 0.583 | 0.365 | 0.413 | 0.217 | 0.461 | 0.266 | 0.315 | 0.154 | 0.364 |
| | dbCAN_2:DIAMOND | 0.688 | 0.759 | 0.221 | 0.830 | 0.737 | 0.785 | 0.219 | 0.834 | 0.743 | 0.812 | 0.217 | 0.881 |
| | dbCAN_2:Hotpep | 0.625 | 0.704 | 0.247 | 0.783 | 0.659 | 0.706 | 0.210 | 0.753 | 0.654 | 0.708 | 0.168 | 0.762 |
| | dbCAN_3 | 0.695 | 0.763 | 0.211 | 0.831 | 0.725 | 0.768 | 0.193 | 0.811 | 0.717 | 0.773 | 0.175 | 0.829 |
| | dbCAN_3:HMMER | 0.437 | 0.509 | 0.227 | 0.582 | 0.366 | 0.413 | 0.215 | 0.461 | 0.269 | 0.318 | 0.153 | 0.367 |
| | dbCAN_3:DIAMOND | 0.739 | 0.802 | 0.198 | 0.865 | 0.829 | 0.867 | 0.173 | 0.906 | 0.896 | 0.933 | 0.114 | 0.969 |
| | dbCAN_3:eCAMI | 0.666 | 0.743 | 0.241 | 0.820 | 0.686 | 0.735 | 0.220 | 0.784 | 0.663 | 0.726 | 0.200 | 0.790 |
| | dbCAN_4 | 0.870 | 0.902 | 0.102 | 0.935 | 0.757 | 0.799 | 0.192 | 0.842 | 0.630 | 0.697 | 0.207 | 0.763 |
| | dbCAN_4:HMMER | 0.535 | 0.609 | 0.229 | 0.682 | 0.422 | 0.474 | 0.236 | 0.527 | 0.291 | 0.339 | 0.153 | 0.388 |
| | dbCAN_4:DIAMOND | 0.732 | 0.795 | 0.199 | 0.859 | 0.840 | 0.877 | 0.170 | 0.915 | 0.937 | 0.959 | 0.069 | 0.981 |
| | dbCAN_4:dbCAN-sub | 0.913 | 0.944 | 0.098 | 0.976 | 0.775 | 0.819 | 0.200 | 0.864 | 0.631 | 0.695 | 0.197 | 0.758 |
| | CUPP | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |

## SI figure 26: The mean sensitivity and 95% confidence interval of the CAZy class classification across all CAZy classes (overleaf)

Sensitivity was calculated for each test set across all CAZy class. The mean sensitivity was then calculated across all test sets and 95% confidence interval (CI) was across all CAZy class, and is plotted in SI figure 26.

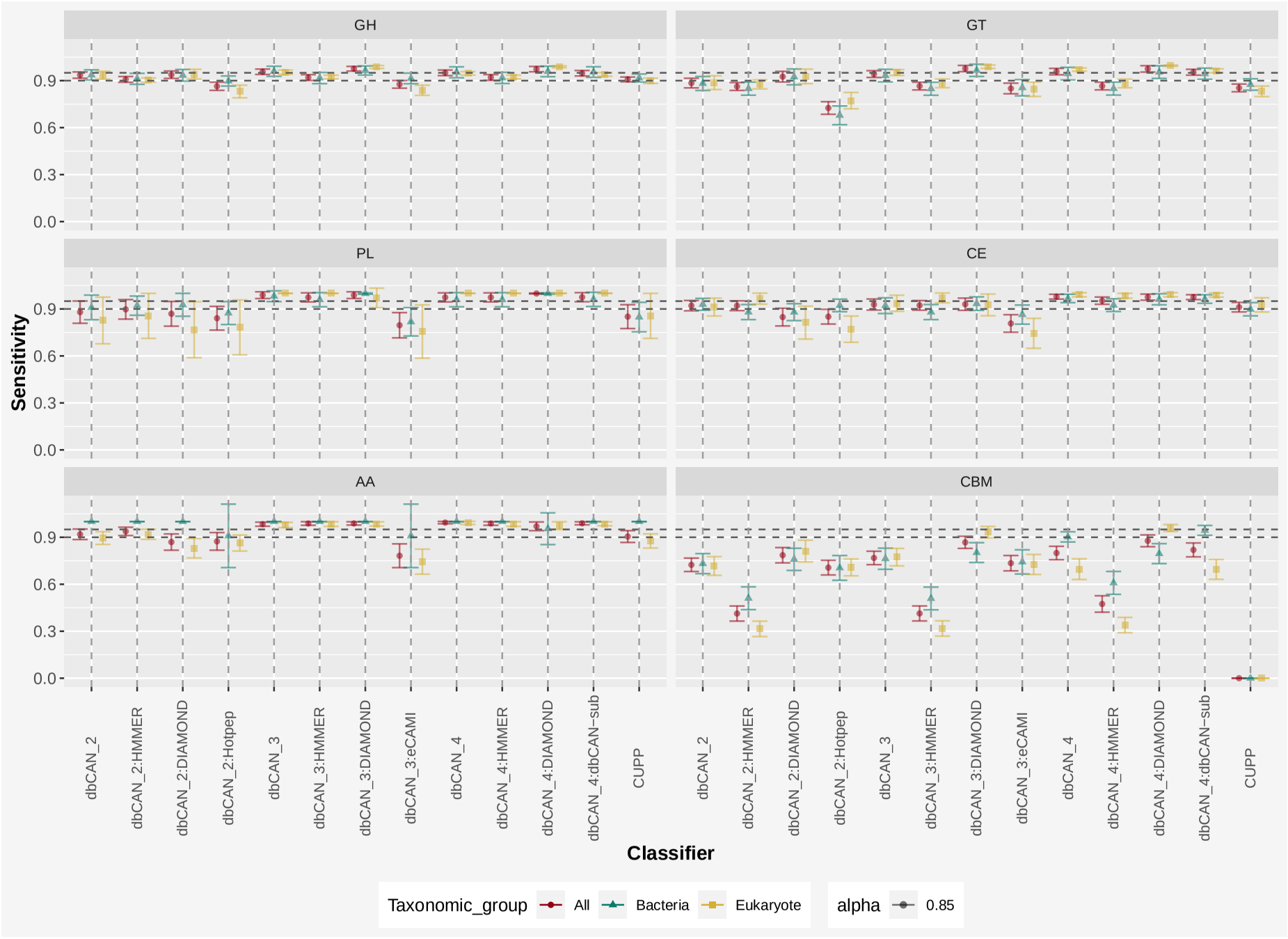



Figure 26: The mean sensitivity and 95% confidence interval (CI) of binary CAZy class classification across all CAZy classes and across all test sets. Where a tool is incorporated into a CAZyme classifier, the tool name is prefixed with the resenstive CAZyme classifier. Per classifier (from left to right) all test sets are pooled together (shaded red); all bacterial test sets are pooled (shaded green); and all test sets from eukaryotic genomes are pooled together (shaded yellow).

**SI figure 27: The sensitivity per test set of the CAZy class classification across all CAZy classes (overleaf)**

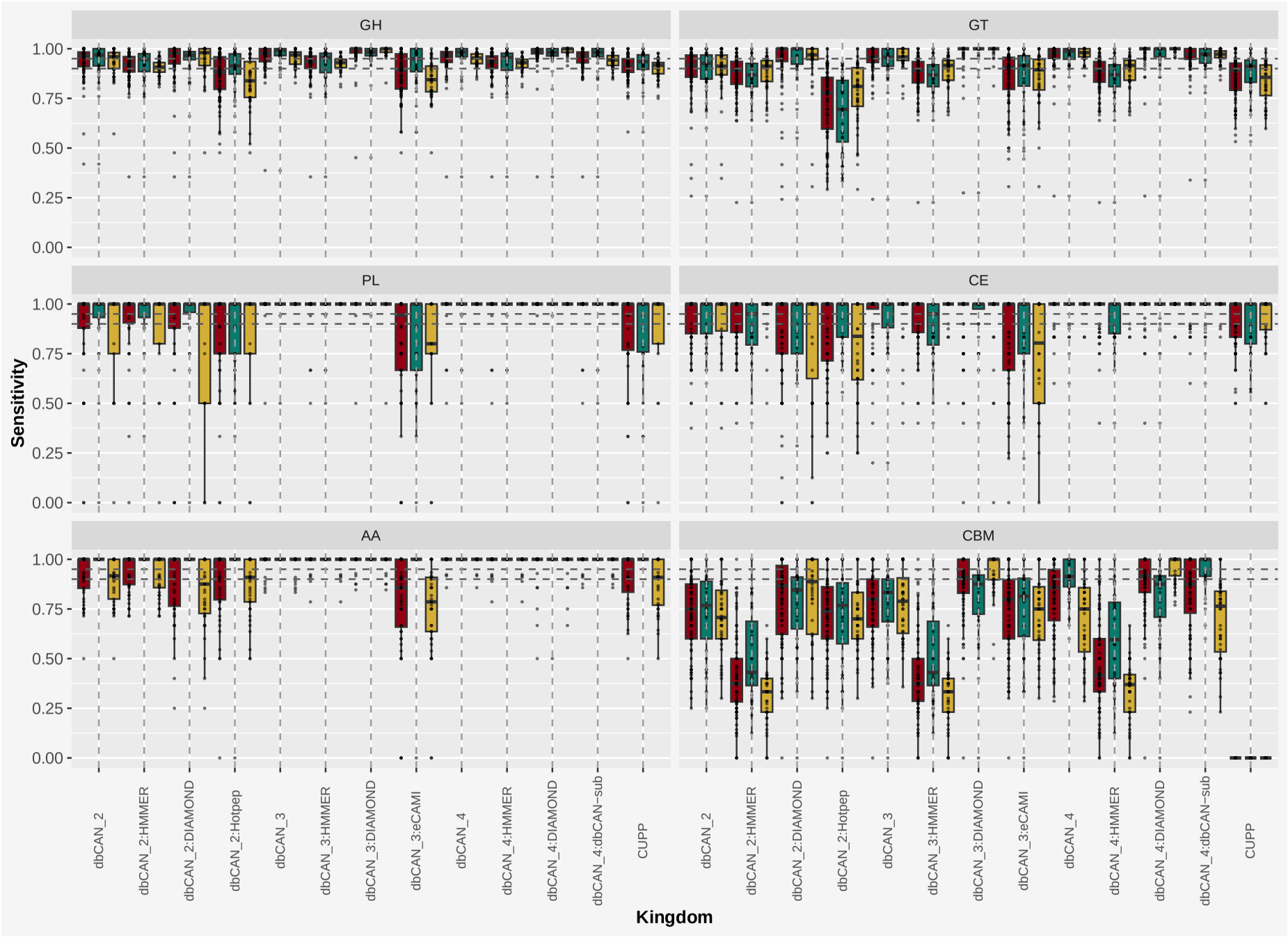



Figure 27: One-dimensional scatter plot overlaying a box and whisker plot, where each point represents the sensitivity of the binary CAZy class classification across all CAZy classes and per test set. Where a tool is incorporated into a CAZyme classifier, the tool name is prefixed with the resenstive CAZyme classifier. Per classifier (from left to right) all test sets are pooled together (shaded red); all bacterial test sets are pooled (shaded green); and all test sets from eukaryotic genomes are pooled together (shaded yellow).

## 8.3 Precision

### SI Table 54: Precision of CAZy class classification per CAZy CAZy class, per taxonomic kingdom (overleaf)

Precision was calculated per test set, the mean (and standard deviation) and 95% confidence interval (CI) was calculated by pooling all test sets.

| CAZy class | Prediction Tool | Kingdom | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | Bacteria | | | | All | | | | Eukaryote | | | |
| | | Lower CI | Mean | SD | Upper CI | Lower CI | Mean | SD | Upper CI | Lower CI | Mean | SD | Upper CI |
| GH | dbCAN_2 | 0.984 | 0.991 | 0.022 | 0.998 | 0.981 | 0.989 | 0.035 | 0.996 | 0.972 | 0.986 | 0.044 | 1.000 |
| | dbCAN_2:HMMER | 0.989 | 0.993 | 0.015 | 0.998 | 0.982 | 0.990 | 0.034 | 0.997 | 0.971 | 0.986 | 0.045 | 1.000 |
| | dbCAN_2:DIAMOND | 0.976 | 0.984 | 0.026 | 0.993 | 0.977 | 0.985 | 0.036 | 0.993 | 0.971 | 0.985 | 0.044 | 0.999 |
| | dbCAN_2:Hotpep | 0.962 | 0.974 | 0.037 | 0.985 | 0.970 | 0.980 | 0.042 | 0.989 | 0.971 | 0.986 | 0.046 | 1.000 |
| | dbCAN_3 | 0.986 | 0.992 | 0.019 | 0.998 | 0.982 | 0.990 | 0.033 | 0.997 | 0.974 | 0.988 | 0.043 | 1.002 |
| | dbCAN_3:HMMER | 0.990 | 0.995 | 0.014 | 0.999 | 0.984 | 0.991 | 0.033 | 0.998 | 0.973 | 0.987 | 0.044 | 1.001 |
| | dbCAN_3:DIAMOND | 0.978 | 0.986 | 0.025 | 0.994 | 0.974 | 0.983 | 0.039 | 0.992 | 0.964 | 0.980 | 0.050 | 0.996 |
| | dbCAN_3:eCAMI | 0.972 | 0.981 | 0.027 | 0.989 | 0.974 | 0.983 | 0.039 | 0.992 | 0.970 | 0.985 | 0.049 | 1.001 |
| | dbCAN_4 | 0.987 | 0.992 | 0.018 | 0.998 | 0.982 | 0.990 | 0.034 | 0.997 | 0.973 | 0.987 | 0.045 | 1.001 |
| | dbCAN_4:HMMER | 0.990 | 0.995 | 0.014 | 0.999 | 0.984 | 0.991 | 0.033 | 0.998 | 0.973 | 0.987 | 0.044 | 1.001 |
| | dbCAN_4:DIAMOND | 0.981 | 0.989 | 0.024 | 0.996 | 0.975 | 0.984 | 0.039 | 0.993 | 0.964 | 0.980 | 0.050 | 0.996 |
| | dbCAN_4:dbCAN-sub | 0.986 | 0.992 | 0.017 | 0.998 | 0.982 | 0.990 | 0.034 | 0.997 | 0.973 | 0.987 | 0.045 | 1.001 |
| | CUPP | 0.989 | 0.993 | 0.014 | 0.998 | 0.983 | 0.991 | 0.034 | 0.998 | 0.973 | 0.988 | 0.046 | 1.003 |
| GT | dbCAN_2 | 0.997 | 0.999 | 0.006 | 1.001 | 0.978 | 0.990 | 0.054 | 1.002 | 0.957 | 0.981 | 0.076 | 1.005 |
| | dbCAN_2:HMMER | 0.997 | 0.999 | 0.005 | 1.001 | 0.976 | 0.988 | 0.057 | 1.001 | 0.953 | 0.978 | 0.079 | 1.003 |
| | dbCAN_2:DIAMOND | 0.996 | 0.998 | 0.007 | 1.001 | 0.976 | 0.989 | 0.057 | 1.001 | 0.954 | 0.979 | 0.079 | 1.004 |
| | dbCAN_2:Hotpep | 0.983 | 0.991 | 0.025 | 0.999 | 0.968 | 0.984 | 0.069 | 0.999 | 0.946 | 0.976 | 0.094 | 1.007 |
| | dbCAN_3 | 0.997 | 0.999 | 0.005 | 1.001 | 0.977 | 0.989 | 0.056 | 1.002 | 0.954 | 0.979 | 0.079 | 1.005 |
| | dbCAN_3:HMMER | 0.997 | 0.999 | 0.005 | 1.001 | 0.976 | 0.988 | 0.057 | 1.001 | 0.953 | 0.978 | 0.079 | 1.003 |
| | dbCAN_3:DIAMOND | 0.995 | 0.998 | 0.009 | 1.000 | 0.970 | 0.984 | 0.062 | 0.998 | 0.943 | 0.970 | 0.086 | 0.998 |
| | dbCAN_3:eCAMI | 0.993 | 0.996 | 0.011 | 1.000 | 0.975 | 0.988 | 0.057 | 1.001 | 0.954 | 0.980 | 0.080 | 1.005 |
| | dbCAN_4 | 0.996 | 0.998 | 0.007 | 1.000 | 0.974 | 0.987 | 0.057 | 0.999 | 0.950 | 0.975 | 0.079 | 1.000 |
| | dbCAN_4:HMMER | 0.997 | 0.999 | 0.005 | 1.001 | 0.976 | 0.988 | 0.057 | 1.001 | 0.953 | 0.978 | 0.079 | 1.003 |
| | dbCAN_4:DIAMOND | 0.995 | 0.997 | 0.009 | 1.000 | 0.970 | 0.984 | 0.064 | 0.998 | 0.942 | 0.971 | 0.088 | 0.999 |
| | dbCAN_4:dbCAN-sub | 0.996 | 0.998 | 0.007 | 1.000 | 0.974 | 0.987 | 0.057 | 0.999 | 0.950 | 0.975 | 0.079 | 1.000 |
| | CUPP | 0.999 | 1.000 | 0.003 | 1.000 | 0.974 | 0.988 | 0.062 | 1.002 | 0.949 | 0.977 | 0.087 | 1.005 |
| PL | dbCAN_2 | 0.898 | 0.967 | 0.183 | 1.035 | 0.893 | 0.953 | 0.205 | 1.013 | 0.804 | 0.929 | 0.244 | 1.055 |
| | dbCAN_2:HMMER | 1.000 | 1.000 | 0.000 | 1.000 | 0.931 | 0.974 | 0.148 | 1.018 | 0.804 | 0.929 | 0.244 | 1.055 |
| | dbCAN_2:DIAMOND | 0.884 | 0.956 | 0.190 | 1.027 | 0.851 | 0.925 | 0.250 | 0.998 | 0.700 | 0.871 | 0.331 | 1.041 |
| | dbCAN_2:Hotpep | 0.990 | 0.996 | 0.015 | 1.002 | 0.890 | 0.951 | 0.205 | 1.011 | 0.700 | 0.871 | 0.331 | 1.041 |
| | dbCAN_3 | 0.994 | 0.998 | 0.011 | 1.002 | 0.962 | 0.985 | 0.077 | 1.007 | 0.896 | 0.961 | 0.125 | 1.025 |
| | dbCAN_3:HMMER | 0.994 | 0.998 | 0.011 | 1.002 | 0.962 | 0.985 | 0.077 | 1.007 | 0.896 | 0.961 | 0.125 | 1.025 |
| | dbCAN_3:DIAMOND | 0.964 | 0.987 | 0.061 | 1.010 | 0.943 | 0.973 | 0.101 | 1.003 | 0.873 | 0.949 | 0.147 | 1.024 |
| | dbCAN_3:eCAMI | 0.902 | 0.968 | 0.180 | 1.034 | 0.862 | 0.933 | 0.245 | 1.005 | 0.700 | 0.871 | 0.331 | 1.041 |
| | dbCAN_4 | 0.994 | 0.998 | 0.011 | 1.002 | 0.962 | 0.985 | 0.077 | 1.007 | 0.896 | 0.961 | 0.125 | 1.025 |
| | dbCAN_4:HMMER | 0.994 | 0.998 | 0.011 | 1.002 | 0.962 | 0.985 | 0.077 | 1.007 | 0.896 | 0.961 | 0.125 | 1.025 |
| | dbCAN_4:DIAMOND | 0.964 | 0.987 | 0.061 | 1.010 | 0.943 | 0.973 | 0.101 | 1.003 | 0.873 | 0.949 | 0.147 | 1.024 |
| | dbCAN_4:dbCAN-sub | 0.994 | 0.998 | 0.011 | 1.002 | 0.962 | 0.985 | 0.077 | 1.007 | 0.896 | 0.961 | 0.125 | 1.025 |
| | CUPP | 0.898 | 0.967 | 0.183 | 1.035 | 0.889 | 0.950 | 0.206 | 1.010 | 0.794 | 0.920 | 0.245 | 1.045 |
| CE | dbCAN_2 | 0.856 | 0.915 | 0.183 | 0.973 | 0.920 | 0.952 | 0.141 | 0.984 | 0.979 | 0.993 | 0.042 | 1.007 |
| | dbCAN_2:HMMER | 0.873 | 0.922 | 0.153 | 0.971 | 0.925 | 0.953 | 0.119 | 0.980 | 0.971 | 0.987 | 0.046 | 1.002 |
| | dbCAN_2:DIAMOND | 0.858 | 0.918 | 0.187 | 0.978 | 0.880 | 0.928 | 0.209 | 0.976 | 0.860 | 0.939 | 0.233 | 1.018 |
| | dbCAN_2:Hotpep | 0.791 | 0.856 | 0.202 | 0.921 | 0.875 | 0.913 | 0.167 | 0.951 | 0.948 | 0.976 | 0.082 | 1.004 |
| | dbCAN_3 | 0.857 | 0.915 | 0.182 | 0.973 | 0.921 | 0.953 | 0.139 | 0.984 | 0.983 | 0.994 | 0.033 | 1.006 |
| | dbCAN_3:HMMER | 0.886 | 0.932 | 0.143 | 0.978 | 0.935 | 0.960 | 0.110 | 0.985 | 0.979 | 0.991 | 0.038 | 1.004 |
| | dbCAN_3:DIAMOND | 0.866 | 0.922 | 0.175 | 0.978 | 0.888 | 0.930 | 0.186 | 0.973 | 0.872 | 0.940 | 0.200 | 1.008 |
| | dbCAN_3:eCAMI | 0.804 | 0.868 | 0.200 | 0.932 | 0.872 | 0.916 | 0.191 | 0.959 | 0.912 | 0.968 | 0.168 | 1.025 |
| | dbCAN_4 | 0.887 | 0.932 | 0.143 | 0.978 | 0.935 | 0.960 | 0.110 | 0.986 | 0.979 | 0.991 | 0.038 | 1.004 |
| | dbCAN_4:HMMER | 0.887 | 0.932 | 0.143 | 0.978 | 0.935 | 0.960 | 0.110 | 0.986 | 0.979 | 0.991 | 0.038 | 1.004 |
| | dbCAN_4:DIAMOND | 0.874 | 0.930 | 0.174 | 0.985 | 0.913 | 0.948 | 0.151 | 0.982 | 0.927 | 0.968 | 0.119 | 1.008 |
| | dbCAN_4:dbCAN-sub | 0.886 | 0.932 | 0.143 | 0.978 | 0.934 | 0.959 | 0.112 | 0.985 | 0.973 | 0.989 | 0.049 | 1.006 |
| | CUPP | 0.883 | 0.930 | 0.145 | 0.976 | 0.934 | 0.960 | 0.113 | 0.986 | 0.979 | 0.993 | 0.042 | 1.007 |
| AA | dbCAN_2 | 1.000 | 1.000 | 0.000 | 1.000 | 0.872 | 0.921 | 0.169 | 0.970 | 0.835 | 0.897 | 0.187 | 0.960 |
| | dbCAN_2:HMMER | 1.000 | 1.000 | 0.000 | 1.000 | 0.872 | 0.919 | 0.161 | 0.966 | 0.836 | 0.895 | 0.176 | 0.954 |
| | dbCAN_2:DIAMOND | 1.000 | 1.000 | 0.000 | 1.000 | 0.873 | 0.923 | 0.171 | 0.972 | 0.837 | 0.900 | 0.189 | 0.963 |
| | dbCAN_2:Hotpep | 0.707 | 0.909 | 0.302 | 1.112 | 0.835 | 0.898 | 0.217 | 0.961 | 0.832 | 0.895 | 0.190 | 0.958 |
| | dbCAN_3 | 1.000 | 1.000 | 0.000 | 1.000 | 0.877 | 0.924 | 0.161 | 0.970 | 0.842 | 0.901 | 0.177 | 0.960 |
| | dbCAN_3:HMMER | 1.000 | 1.000 | 0.000 | 1.000 | 0.872 | 0.919 | 0.162 | 0.966 | 0.836 | 0.895 | 0.179 | 0.955 |
| | dbCAN_3:DIAMOND | 1.000 | 1.000 | 0.000 | 1.000 | 0.868 | 0.917 | 0.169 | 0.966 | 0.830 | 0.892 | 0.186 | 0.955 |
| | dbCAN_3:eCAMI | 0.707 | 0.909 | 0.302 | 1.112 | 0.770 | 0.851 | 0.281 | 0.933 | 0.742 | 0.834 | 0.276 | 0.926 |
| | dbCAN_4 | 1.000 | 1.000 | 0.000 | 1.000 | 0.873 | 0.920 | 0.161 | 0.967 | 0.837 | 0.896 | 0.177 | 0.955 |
| | dbCAN_4:HMMER | 1.000 | 1.000 | 0.000 | 1.000 | 0.872 | 0.919 | 0.162 | 0.966 | 0.836 | 0.895 | 0.179 | 0.955 |
| | dbCAN_4:DIAMOND | 1.000 | 1.000 | 0.000 | 1.000 | 0.870 | 0.920 | 0.170 | 0.969 | 0.833 | 0.896 | 0.187 | 0.958 |
| | dbCAN_4:dbCAN-sub | 1.000 | 1.000 | 0.000 | 1.000 | 0.871 | 0.918 | 0.162 | 0.965 | 0.834 | 0.893 | 0.178 | 0.952 |
| | CUPP | 1.000 | 1.000 | 0.000 | 1.000 | 0.869 | 0.918 | 0.170 | 0.968 | 0.832 | 0.894 | 0.187 | 0.957 |
| CBM | dbCAN_2 | 0.917 | 0.952 | 0.109 | 0.987 | 0.891 | 0.920 | 0.133 | 0.950 | 0.841 | 0.889 | 0.149 | 0.936 |
| | dbCAN_2:HMMER | 0.894 | 0.938 | 0.136 | 0.982 | 0.845 | 0.897 | 0.230 | 0.948 | 0.762 | 0.855 | 0.292 | 0.948 |
| | dbCAN_2:DIAMOND | 0.903 | 0.955 | 0.163 | 1.007 | 0.889 | 0.922 | 0.150 | 0.955 | 0.847 | 0.889 | 0.130 | 0.930 |
| | dbCAN_2:Hotpep | 0.416 | 0.480 | 0.200 | 0.544 | 0.431 | 0.470 | 0.173 | 0.508 | 0.414 | 0.460 | 0.143 | 0.505 |
| | dbCAN_3 | 0.884 | 0.938 | 0.170 | 0.993 | 0.913 | 0.943 | 0.136 | 0.973 | 0.919 | 0.948 | 0.090 | 0.977 |
| | dbCAN_3:HMMER | 0.894 | 0.938 | 0.137 | 0.981 | 0.845 | 0.896 | 0.230 | 0.948 | 0.762 | 0.855 | 0.292 | 0.948 |
| | dbCAN_3:DIAMOND | 0.909 | 0.960 | 0.159 | 1.011 | 0.923 | 0.951 | 0.128 | 0.980 | 0.914 | 0.942 | 0.089 | 0.971 |
| | dbCAN_3:eCAMI | 0.558 | 0.632 | 0.231 | 0.705 | 0.611 | 0.662 | 0.225 | 0.712 | 0.621 | 0.691 | 0.219 | 0.761 |
| | dbCAN_4 | 0.949 | 0.970 | 0.066 | 0.991 | 0.936 | 0.956 | 0.092 | 0.977 | 0.907 | 0.943 | 0.111 | 0.978 |
| | dbCAN_4:HMMER | 0.916 | 0.952 | 0.113 | 0.989 | 0.874 | 0.918 | 0.200 | 0.963 | 0.802 | 0.884 | 0.257 | 0.966 |
| | dbCAN_4:DIAMOND | 0.910 | 0.960 | 0.159 | 1.011 | 0.929 | 0.957 | 0.127 | 0.985 | 0.926 | 0.954 | 0.087 | 0.981 |
| | dbCAN_4:dbCAN-sub | 0.941 | 0.962 | 0.068 | 0.984 | 0.912 | 0.938 | 0.115 | 0.963 | 0.867 | 0.913 | 0.145 | 0.959 |
| | CUPP | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |

## SI figure 28: The mean precision and 95% confidence interval of the CAZy class classification across all CAZy classes (overleaf)

Precision was calculated for each test set across all CAZy class. The mean precision was then calculated across all test sets and 95% confidence interval (CI) was across all CAZy class, and is plotted in SI figure 28.

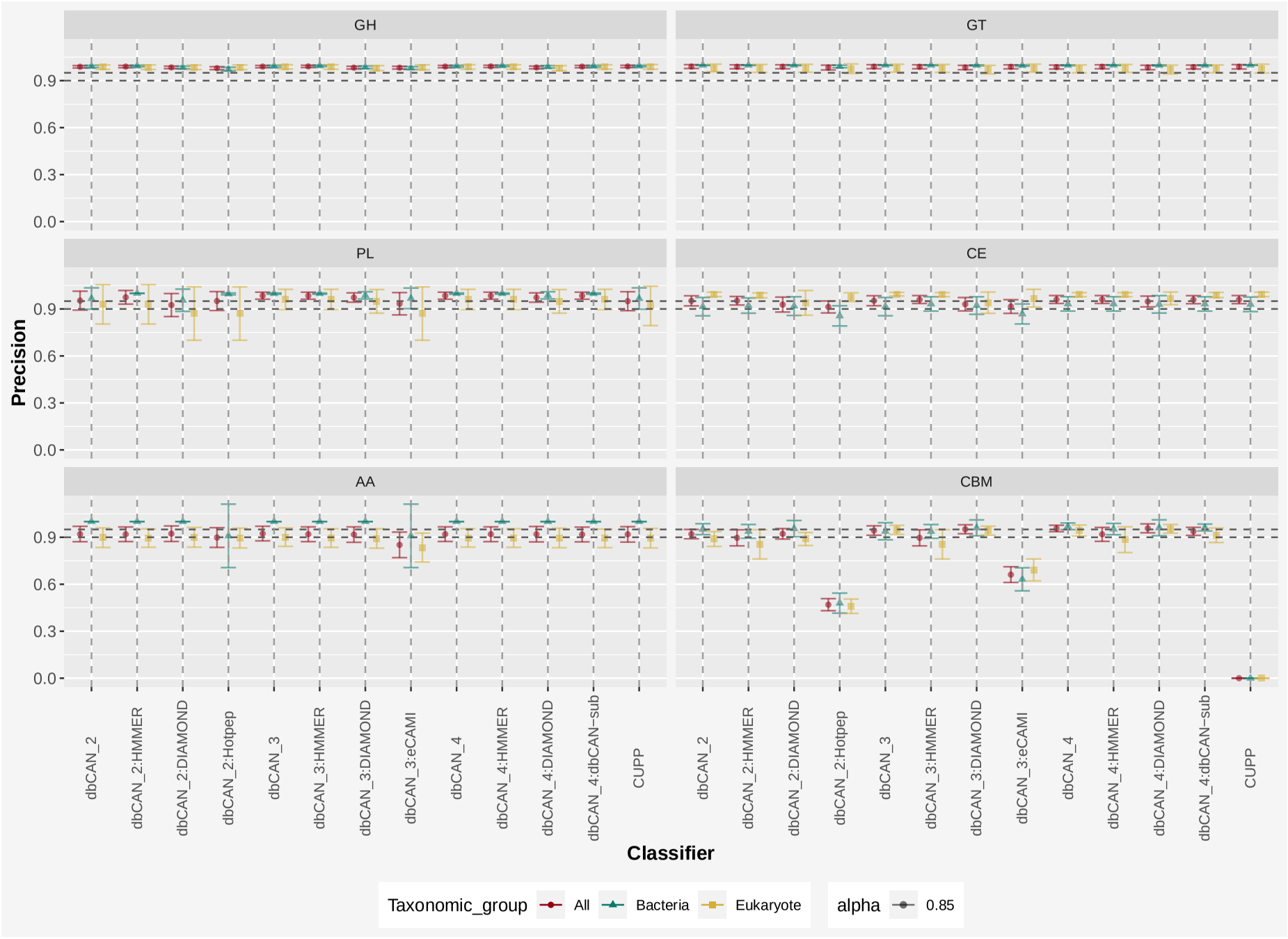



Figure 28: The mean precision and 95% confidence interval (CI) of binary CAZy class classification across all CAZy classes and across all test sets. Where a tool is incorporated into a CAZyme classifier, the tool name is prefixed with the reprective CAZyme classifier. Per classifier (from left to right) all test sets are pooled together (shaded red); all bacterial test sets are pooled (shaded green); and all test sets from eukaryotic genomes are pooled together (shaded yellow).

**SI figure 29: The precision per test set of the CAZy class classification across all CAZy classes (overleaf)**

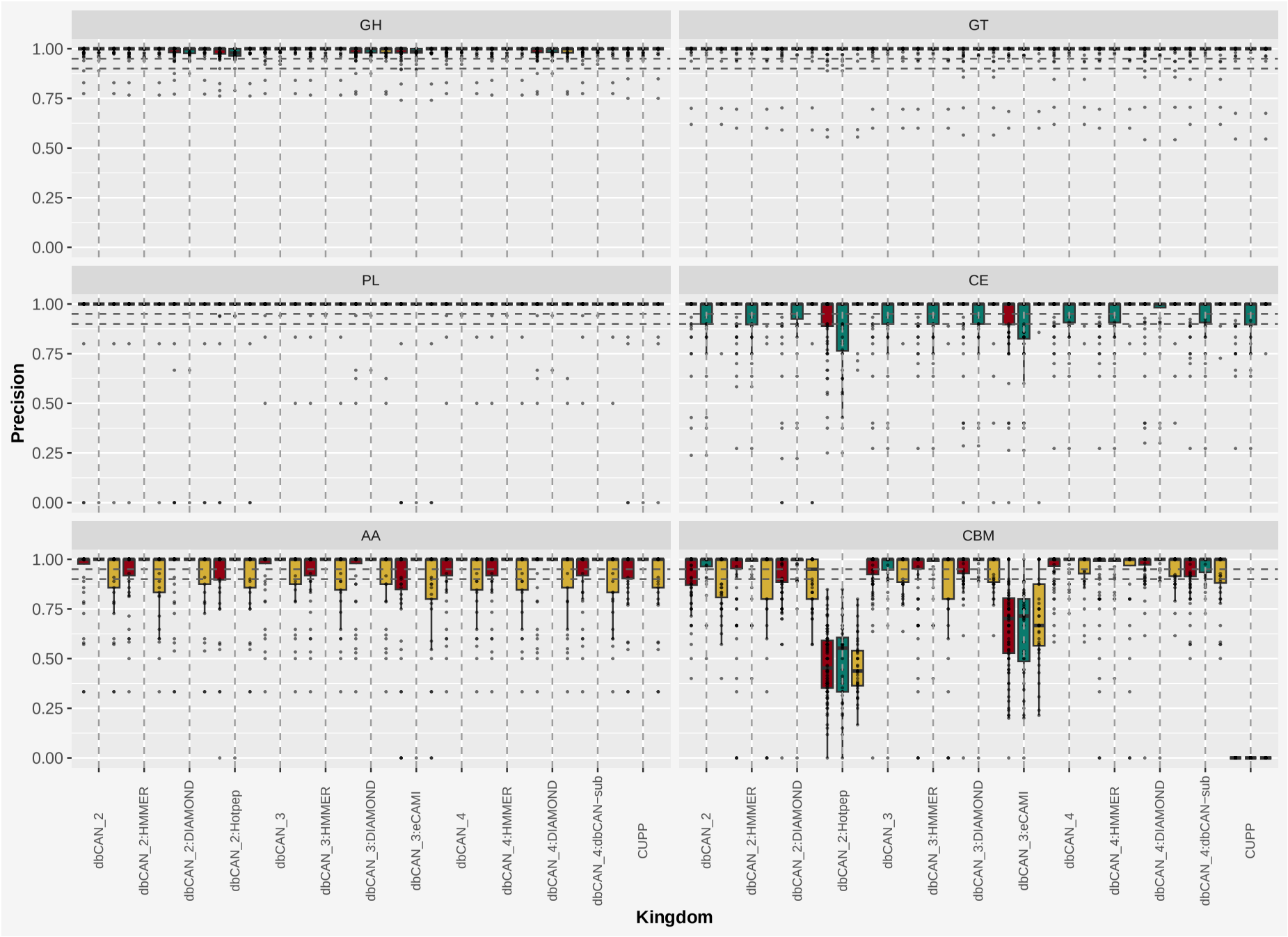



Figure 29: One-dimensional scatter plot overlaying a box and whisker plot, where each point represents the precision of the binary CAZy class classification across all CAZy classes and per test set. Where a tool is incorporated into a CAZyme classifier, the tool name is prefixed with the reprective CAZyme classifier. Per classifier (from left to right) all test sets are pooled together (shaded red); all bacterial test sets are pooled (shaded green); and all test sets from eukaryotic genomes are pooled together (shaded yellow).

## 8.4 F1-score

### SI Table 55: F1-score of CAZy class classification per CAZy CAZy class, per taxonomic kingdom (overleaf)

F1-score was calculated per test set, the mean (and standard deviation) and 95% confidence interval (CI) was calculated by pooling all test sets.

| CAZy class | Prediction Tool | Kingdom | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | Bacteria | | | | All | | | | Eukaryote | | | |
| | | Lower CI | Mean | SD | Upper CI | Lower CI | Mean | SD | Upper CI | Lower CI | Mean | SD | Upper CI |
| GH | dbCAN_2 | 0.937 | 0.959 | 0.069 | 0.981 | 0.945 | 0.958 | 0.061 | 0.972 | 0.940 | 0.957 | 0.053 | 0.974 |
| | dbCAN_2:HMMER | 0.922 | 0.946 | 0.077 | 0.971 | 0.931 | 0.944 | 0.059 | 0.957 | 0.932 | 0.942 | 0.033 | 0.953 |
| | dbCAN_2:DIAMOND | 0.928 | 0.954 | 0.081 | 0.979 | 0.940 | 0.956 | 0.073 | 0.972 | 0.938 | 0.959 | 0.064 | 0.980 |
| | dbCAN_2:Hotpep | 0.911 | 0.931 | 0.063 | 0.951 | 0.896 | 0.914 | 0.077 | 0.931 | 0.869 | 0.896 | 0.087 | 0.924 |
| | dbCAN_3 | 0.949 | 0.972 | 0.071 | 0.995 | 0.958 | 0.971 | 0.056 | 0.983 | 0.958 | 0.970 | 0.035 | 0.981 |
| | dbCAN_3:HMMER | 0.924 | 0.949 | 0.078 | 0.974 | 0.938 | 0.951 | 0.060 | 0.965 | 0.943 | 0.953 | 0.033 | 0.964 |
| | dbCAN_3:DIAMOND | 0.953 | 0.972 | 0.061 | 0.992 | 0.967 | 0.978 | 0.049 | 0.989 | 0.972 | 0.983 | 0.033 | 0.993 |
| | dbCAN_3:eCAMI | 0.922 | 0.942 | 0.064 | 0.963 | 0.907 | 0.922 | 0.070 | 0.938 | 0.880 | 0.902 | 0.070 | 0.925 |
| | dbCAN_4 | 0.943 | 0.968 | 0.078 | 0.993 | 0.954 | 0.967 | 0.059 | 0.980 | 0.956 | 0.966 | 0.032 | 0.976 |
| | dbCAN_4:HMMER | 0.925 | 0.950 | 0.078 | 0.975 | 0.938 | 0.952 | 0.060 | 0.965 | 0.943 | 0.953 | 0.033 | 0.964 |
| | dbCAN_4:DIAMOND | 0.945 | 0.969 | 0.076 | 0.994 | 0.963 | 0.976 | 0.058 | 0.989 | 0.974 | 0.983 | 0.031 | 0.993 |
| | dbCAN_4:dbCAN-sub | 0.945 | 0.969 | 0.076 | 0.993 | 0.953 | 0.966 | 0.058 | 0.978 | 0.952 | 0.962 | 0.032 | 0.972 |
| | CUPP | 0.936 | 0.952 | 0.049 | 0.967 | 0.936 | 0.946 | 0.045 | 0.956 | 0.927 | 0.941 | 0.042 | 0.954 |
| GT | dbCAN_2 | 0.897 | 0.929 | 0.102 | 0.962 | 0.903 | 0.926 | 0.102 | 0.948 | 0.890 | 0.922 | 0.102 | 0.955 |
| | dbCAN_2:HMMER | 0.879 | 0.911 | 0.101 | 0.943 | 0.896 | 0.915 | 0.085 | 0.934 | 0.899 | 0.919 | 0.065 | 0.940 |
| | dbCAN_2:DIAMOND | 0.914 | 0.951 | 0.115 | 0.987 | 0.921 | 0.946 | 0.112 | 0.971 | 0.906 | 0.942 | 0.111 | 0.978 |
| | dbCAN_2:Hotpep | 0.745 | 0.790 | 0.142 | 0.836 | 0.790 | 0.821 | 0.138 | 0.852 | 0.810 | 0.851 | 0.129 | 0.893 |
| | dbCAN_3 | 0.927 | 0.958 | 0.097 | 0.989 | 0.943 | 0.961 | 0.079 | 0.978 | 0.945 | 0.963 | 0.057 | 0.981 |
| | dbCAN_3:HMMER | 0.878 | 0.911 | 0.101 | 0.943 | 0.898 | 0.917 | 0.084 | 0.936 | 0.903 | 0.923 | 0.064 | 0.943 |
| | dbCAN_3:DIAMOND | 0.946 | 0.976 | 0.092 | 1.005 | 0.960 | 0.976 | 0.076 | 0.993 | 0.959 | 0.977 | 0.055 | 0.994 |
| | dbCAN_3:eCAMI | 0.877 | 0.911 | 0.108 | 0.946 | 0.881 | 0.905 | 0.108 | 0.929 | 0.863 | 0.898 | 0.109 | 0.933 |
| | dbCAN_4 | 0.934 | 0.965 | 0.096 | 0.995 | 0.951 | 0.968 | 0.075 | 0.985 | 0.956 | 0.971 | 0.048 | 0.986 |
| | dbCAN_4:HMMER | 0.879 | 0.911 | 0.101 | 0.943 | 0.898 | 0.917 | 0.084 | 0.936 | 0.903 | 0.923 | 0.063 | 0.943 |
| | dbCAN_4:DIAMOND | 0.939 | 0.970 | 0.096 | 1.000 | 0.958 | 0.975 | 0.078 | 0.992 | 0.963 | 0.981 | 0.055 | 0.998 |
| | dbCAN_4:dbCAN-sub | 0.940 | 0.966 | 0.080 | 0.992 | 0.952 | 0.966 | 0.066 | 0.981 | 0.951 | 0.967 | 0.048 | 0.982 |
| | CUPP | 0.906 | 0.929 | 0.071 | 0.952 | 0.894 | 0.911 | 0.076 | 0.928 | 0.868 | 0.892 | 0.077 | 0.917 |
| PL | dbCAN_2 | 0.861 | 0.933 | 0.192 | 1.005 | 0.843 | 0.907 | 0.218 | 0.971 | 0.729 | 0.862 | 0.259 | 0.995 |
| | dbCAN_2:HMMER | 0.907 | 0.950 | 0.114 | 0.992 | 0.873 | 0.925 | 0.179 | 0.978 | 0.750 | 0.882 | 0.256 | 1.013 |
| | dbCAN_2:DIAMOND | 0.868 | 0.937 | 0.187 | 1.007 | 0.815 | 0.889 | 0.253 | 0.963 | 0.634 | 0.803 | 0.329 | 0.973 |
| | dbCAN_2:Hotpep | 0.868 | 0.917 | 0.133 | 0.967 | 0.814 | 0.880 | 0.226 | 0.947 | 0.646 | 0.815 | 0.328 | 0.983 |
| | dbCAN_3 | 0.964 | 0.987 | 0.061 | 1.010 | 0.962 | 0.983 | 0.069 | 1.003 | 0.933 | 0.975 | 0.082 | 1.017 |
| | dbCAN_3:HMMER | 0.945 | 0.974 | 0.077 | 1.002 | 0.951 | 0.974 | 0.078 | 0.997 | 0.933 | 0.975 | 0.082 | 1.017 |
| | dbCAN_3:DIAMOND | 0.978 | 0.991 | 0.037 | 1.005 | 0.952 | 0.975 | 0.079 | 0.999 | 0.886 | 0.947 | 0.119 | 1.009 |
| | dbCAN_3:eCAMI | 0.797 | 0.873 | 0.208 | 0.950 | 0.774 | 0.847 | 0.253 | 0.921 | 0.634 | 0.800 | 0.321 | 0.965 |
| | dbCAN_4 | 0.945 | 0.974 | 0.077 | 1.002 | 0.951 | 0.974 | 0.078 | 0.997 | 0.933 | 0.975 | 0.082 | 1.017 |
| | dbCAN_4:HMMER | 0.945 | 0.974 | 0.077 | 1.002 | 0.951 | 0.974 | 0.078 | 0.997 | 0.933 | 0.975 | 0.082 | 1.017 |
| | dbCAN_4:DIAMOND | 0.978 | 0.991 | 0.037 | 1.005 | 0.963 | 0.982 | 0.065 | 1.001 | 0.918 | 0.967 | 0.095 | 1.016 |
| | dbCAN_4:dbCAN-sub | 0.946 | 0.975 | 0.077 | 1.003 | 0.952 | 0.975 | 0.078 | 0.998 | 0.933 | 0.975 | 0.082 | 1.017 |
| | CUPP | 0.809 | 0.890 | 0.218 | 0.971 | 0.818 | 0.885 | 0.229 | 0.952 | 0.746 | 0.876 | 0.254 | 1.007 |
| CE | dbCAN_2 | 0.860 | 0.905 | 0.141 | 0.950 | 0.893 | 0.922 | 0.130 | 0.952 | 0.902 | 0.941 | 0.116 | 0.981 |
| | dbCAN_2:HMMER | 0.842 | 0.882 | 0.126 | 0.923 | 0.899 | 0.925 | 0.112 | 0.950 | 0.949 | 0.972 | 0.068 | 0.995 |
| | dbCAN_2:DIAMOND | 0.824 | 0.876 | 0.163 | 0.928 | 0.813 | 0.864 | 0.223 | 0.915 | 0.756 | 0.850 | 0.278 | 0.944 |
| | dbCAN_2:Hotpep | 0.822 | 0.873 | 0.158 | 0.924 | 0.816 | 0.855 | 0.171 | 0.894 | 0.774 | 0.836 | 0.184 | 0.898 |
| | dbCAN_3 | 0.842 | 0.894 | 0.162 | 0.946 | 0.892 | 0.924 | 0.140 | 0.956 | 0.922 | 0.956 | 0.102 | 0.991 |
| | dbCAN_3:HMMER | 0.849 | 0.888 | 0.122 | 0.926 | 0.906 | 0.930 | 0.107 | 0.955 | 0.957 | 0.978 | 0.062 | 0.998 |
| | dbCAN_3:DIAMOND | 0.863 | 0.909 | 0.144 | 0.955 | 0.877 | 0.915 | 0.168 | 0.954 | 0.857 | 0.922 | 0.193 | 0.987 |
| | dbCAN_3:eCAMI | 0.787 | 0.844 | 0.178 | 0.901 | 0.785 | 0.831 | 0.202 | 0.878 | 0.740 | 0.817 | 0.227 | 0.894 |
| | dbCAN_4 | 0.905 | 0.940 | 0.110 | 0.975 | 0.944 | 0.964 | 0.085 | 0.983 | 0.981 | 0.990 | 0.029 | 1.000 |
| | dbCAN_4:HMMER | 0.876 | 0.914 | 0.119 | 0.952 | 0.927 | 0.949 | 0.096 | 0.971 | 0.976 | 0.987 | 0.031 | 0.997 |
| | dbCAN_4:DIAMOND | 0.884 | 0.930 | 0.143 | 0.975 | 0.923 | 0.951 | 0.119 | 0.978 | 0.946 | 0.974 | 0.082 | 1.002 |
| | dbCAN_4:dbCAN-sub | 0.903 | 0.938 | 0.110 | 0.973 | 0.942 | 0.962 | 0.086 | 0.981 | 0.977 | 0.988 | 0.034 | 0.999 |
| | CUPP | 0.862 | 0.900 | 0.120 | 0.938 | 0.900 | 0.925 | 0.109 | 0.950 | 0.922 | 0.953 | 0.091 | 0.984 |
| AA | dbCAN_2 | 1.000 | 1.000 | 0.000 | 1.000 | 0.872 | 0.910 | 0.132 | 0.949 | 0.837 | 0.884 | 0.140 | 0.930 |
| | dbCAN_2:HMMER | 1.000 | 1.000 | 0.000 | 1.000 | 0.884 | 0.917 | 0.114 | 0.950 | 0.853 | 0.893 | 0.119 | 0.932 |
| | dbCAN_2:DIAMOND | 1.000 | 1.000 | 0.000 | 1.000 | 0.830 | 0.877 | 0.160 | 0.923 | 0.785 | 0.840 | 0.166 | 0.895 |
| | dbCAN_2:Hotpep | 0.707 | 0.909 | 0.302 | 1.112 | 0.817 | 0.873 | 0.192 | 0.929 | 0.813 | 0.863 | 0.150 | 0.913 |
| | dbCAN_3 | 1.000 | 1.000 | 0.000 | 1.000 | 0.912 | 0.944 | 0.110 | 0.976 | 0.887 | 0.927 | 0.121 | 0.967 |
| | dbCAN_3:HMMER | 1.000 | 1.000 | 0.000 | 1.000 | 0.911 | 0.943 | 0.111 | 0.976 | 0.886 | 0.927 | 0.122 | 0.967 |
| | dbCAN_3:DIAMOND | 1.000 | 1.000 | 0.000 | 1.000 | 0.908 | 0.941 | 0.114 | 0.974 | 0.882 | 0.924 | 0.125 | 0.966 |
| | dbCAN_3:eCAMI | 0.707 | 0.909 | 0.302 | 1.112 | 0.726 | 0.801 | 0.256 | 0.875 | 0.690 | 0.768 | 0.236 | 0.847 |
| | dbCAN_4 | 1.000 | 1.000 | 0.000 | 1.000 | 0.915 | 0.946 | 0.109 | 0.978 | 0.891 | 0.931 | 0.119 | 0.970 |
| | dbCAN_4:HMMER | 1.000 | 1.000 | 0.000 | 1.000 | 0.911 | 0.943 | 0.111 | 0.976 | 0.886 | 0.927 | 0.122 | 0.967 |
| | dbCAN_4:DIAMOND | 0.902 | 0.970 | 0.101 | 1.037 | 0.895 | 0.931 | 0.123 | 0.966 | 0.877 | 0.919 | 0.127 | 0.961 |
| | dbCAN_4:dbCAN-sub | 1.000 | 1.000 | 0.000 | 1.000 | 0.911 | 0.943 | 0.109 | 0.975 | 0.886 | 0.926 | 0.119 | 0.966 |
| | CUPP | 1.000 | 1.000 | 0.000 | 1.000 | 0.861 | 0.901 | 0.138 | 0.941 | 0.824 | 0.872 | 0.144 | 0.920 |
| CBM | dbCAN_2 | 0.763 | 0.814 | 0.160 | 0.866 | 0.763 | 0.797 | 0.154 | 0.831 | 0.732 | 0.780 | 0.148 | 0.827 |
| | dbCAN_2:HMMER | 0.571 | 0.635 | 0.199 | 0.698 | 0.494 | 0.541 | 0.215 | 0.589 | 0.387 | 0.448 | 0.190 | 0.509 |
| | dbCAN_2:DIAMOND | 0.773 | 0.833 | 0.187 | 0.893 | 0.793 | 0.833 | 0.176 | 0.872 | 0.779 | 0.832 | 0.167 | 0.886 |
| | dbCAN_2:Hotpep | 0.492 | 0.558 | 0.206 | 0.624 | 0.512 | 0.551 | 0.172 | 0.589 | 0.501 | 0.543 | 0.132 | 0.586 |
| | dbCAN_3 | 0.773 | 0.833 | 0.186 | 0.892 | 0.801 | 0.835 | 0.155 | 0.870 | 0.800 | 0.838 | 0.117 | 0.875 |
| | dbCAN_3:HMMER | 0.570 | 0.634 | 0.198 | 0.697 | 0.495 | 0.542 | 0.213 | 0.590 | 0.391 | 0.451 | 0.190 | 0.512 |
| | dbCAN_3:DIAMOND | 0.811 | 0.865 | 0.169 | 0.919 | 0.868 | 0.899 | 0.138 | 0.929 | 0.904 | 0.932 | 0.088 | 0.961 |
| | dbCAN_3:eCAMI | 0.599 | 0.671 | 0.225 | 0.743 | 0.635 | 0.681 | 0.208 | 0.727 | 0.630 | 0.691 | 0.192 | 0.752 |
| | dbCAN_4 | 0.910 | 0.930 | 0.060 | 0.949 | 0.824 | 0.855 | 0.139 | 0.886 | 0.730 | 0.780 | 0.155 | 0.829 |
| | dbCAN_4:HMMER | 0.657 | 0.716 | 0.185 | 0.775 | 0.546 | 0.595 | 0.220 | 0.644 | 0.416 | 0.474 | 0.184 | 0.533 |
| | dbCAN_4:DIAMOND | 0.806 | 0.861 | 0.170 | 0.915 | 0.877 | 0.907 | 0.136 | 0.937 | 0.933 | 0.954 | 0.066 | 0.975 |
| | dbCAN_4:dbCAN-sub | 0.928 | 0.948 | 0.063 | 0.968 | 0.824 | 0.858 | 0.149 | 0.891 | 0.717 | 0.767 | 0.156 | 0.817 |
| | CUPP | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |

## SI figure 30: The mean F1-score and 95% confidence interval of the CAZy class classification across all CAZy classes (overleaf)

F1-score was calculated for each test set across all CAZy class. The mean F1-score was then calculated across all test sets and 95% confidence interval (CI) was across all CAZy class, and is plotted in SI figure 30.

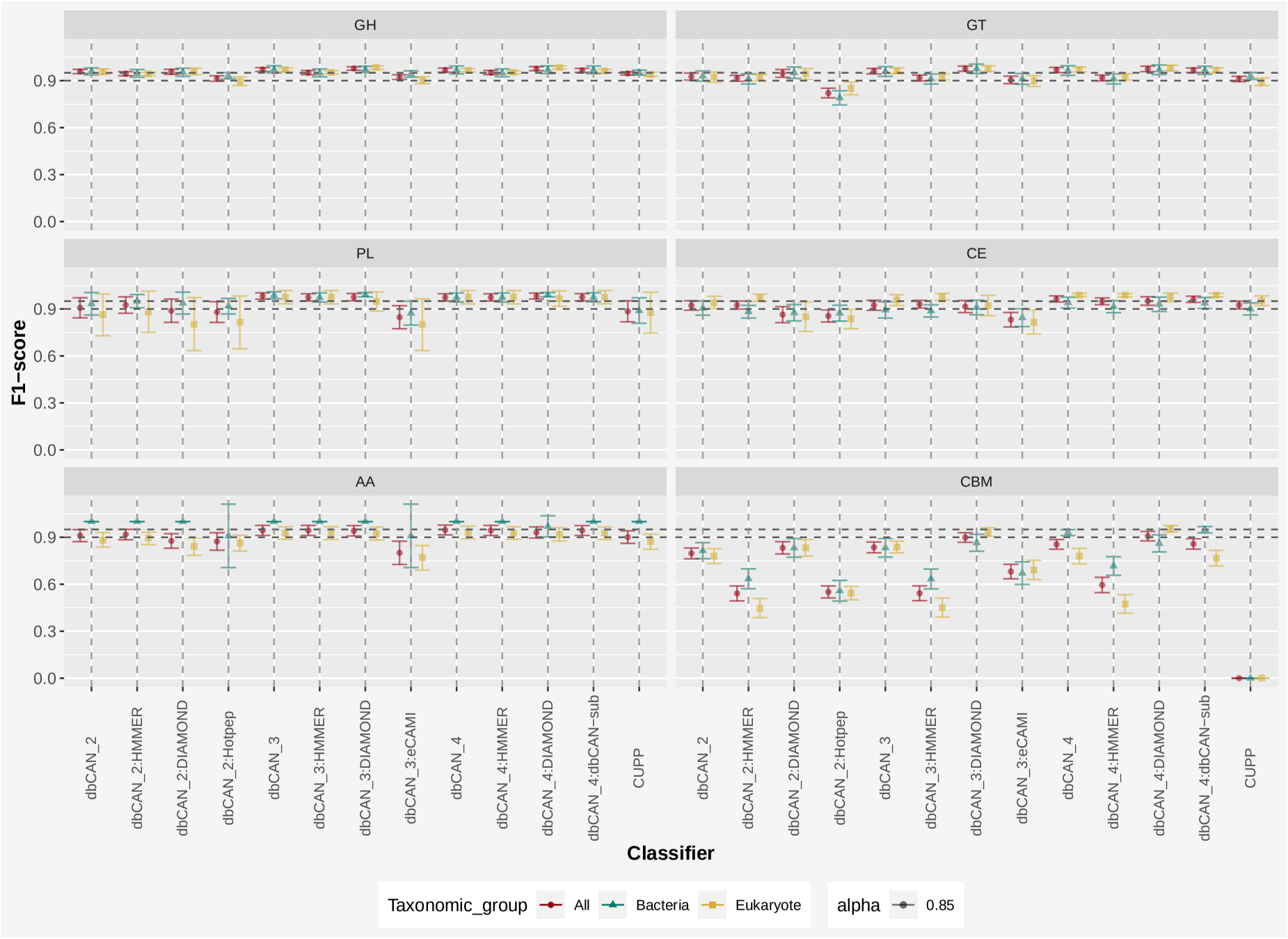



Figure 30: The mean F1-score and 95% confidence interval (CI) of binary CAZy class classification across all CAZy classes and across all test sets. Where a tool is incorporated into a CAZyme classifier, the tool name is prefixed with the ref1tive CAZyme classifier. Per classifier (from left to right) all test sets are pooled together (shaded red); all bacterial test sets are pooled (shaded green); and all test sets from eukaryotic genomes are pooled together (shaded yellow).

**SI figure 31: The F1-score per test set of the CAZy class classification across all CAZy classes (overleaf)**

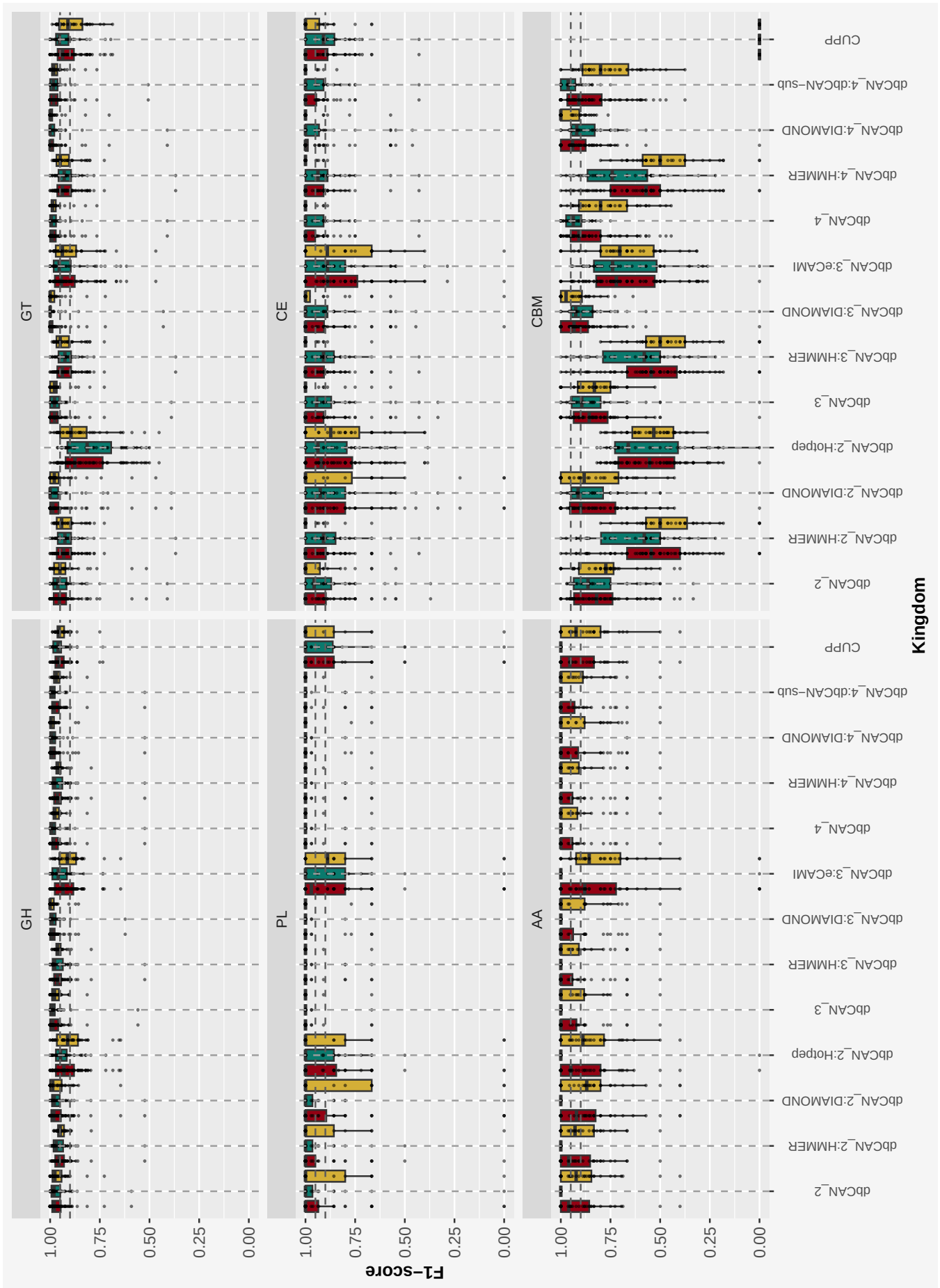



Figure 31: One-dimensional scatter plot overlaying a box and whisker plot, where each point represents the F1-score of the binary CAZy class classification across all CAZy classes and per test set. Where a tool is incorporated into a CAZyme classifier, the tool name is prefixed with the ref1tive CAZyme classifier. Per classifier (from left to right) all test sets are pooled together (shaded red); all bacterial test sets are pooled (shaded green); and all test sets from eukaryotic genomes are pooled together (shaded yellow).

## 8.5 Accuracy

### SI Table 56: Accuracy of CAZy class classification per CAZy CAZy class, per taxonomic kingdom (overleaf)

Accuracy was calculated per test set, the mean (and standard deviation) and 95% confidence interval (CI) was calculated by pooling all test sets.

| CAZy class | Prediction Tool | Kingdom | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | Bacteria | | | | All | | | | Eukaryote | | | |
| | | Lower CI | Mean | SD | Upper CI | Lower CI | Mean | SD | Upper CI | Lower CI | Mean | SD | Upper CI |
| GH | dbCAN_2 | 0.954 | 0.966 | 0.039 | 0.978 | 0.958 | 0.966 | 0.036 | 0.974 | 0.955 | 0.966 | 0.034 | 0.976 |
| | dbCAN_2:HMMER | 0.942 | 0.954 | 0.040 | 0.967 | 0.946 | 0.954 | 0.033 | 0.961 | 0.945 | 0.953 | 0.025 | 0.961 |
| | dbCAN_2:DIAMOND | 0.947 | 0.961 | 0.044 | 0.975 | 0.955 | 0.964 | 0.041 | 0.974 | 0.955 | 0.968 | 0.039 | 0.980 |
| | dbCAN_2:Hotpep | 0.923 | 0.937 | 0.045 | 0.952 | 0.919 | 0.930 | 0.049 | 0.941 | 0.905 | 0.922 | 0.053 | 0.939 |
| | dbCAN_3 | 0.968 | 0.979 | 0.035 | 0.990 | 0.970 | 0.977 | 0.030 | 0.984 | 0.967 | 0.975 | 0.026 | 0.983 |
| | dbCAN_3:HMMER | 0.945 | 0.958 | 0.041 | 0.971 | 0.953 | 0.960 | 0.033 | 0.967 | 0.954 | 0.962 | 0.023 | 0.969 |
| | dbCAN_3:DIAMOND | 0.967 | 0.977 | 0.032 | 0.988 | 0.975 | 0.981 | 0.029 | 0.988 | 0.977 | 0.985 | 0.026 | 0.993 |
| | dbCAN_3:eCAMI | 0.931 | 0.947 | 0.050 | 0.963 | 0.926 | 0.936 | 0.048 | 0.947 | 0.911 | 0.926 | 0.045 | 0.940 |
| | dbCAN_4 | 0.964 | 0.976 | 0.038 | 0.988 | 0.968 | 0.974 | 0.031 | 0.981 | 0.966 | 0.973 | 0.022 | 0.979 |
| | dbCAN_4:HMMER | 0.945 | 0.958 | 0.040 | 0.971 | 0.953 | 0.960 | 0.033 | 0.967 | 0.954 | 0.962 | 0.023 | 0.969 |
| | dbCAN_4:DIAMOND | 0.965 | 0.977 | 0.036 | 0.988 | 0.974 | 0.981 | 0.030 | 0.988 | 0.979 | 0.986 | 0.023 | 0.993 |
| | dbCAN_4:dbCAN-sub | 0.965 | 0.976 | 0.036 | 0.988 | 0.967 | 0.973 | 0.030 | 0.980 | 0.963 | 0.970 | 0.021 | 0.977 |
| | CUPP | 0.945 | 0.955 | 0.031 | 0.965 | 0.947 | 0.954 | 0.029 | 0.961 | 0.944 | 0.952 | 0.027 | 0.961 |
| GT | dbCAN_2 | 0.934 | 0.958 | 0.075 | 0.982 | 0.938 | 0.955 | 0.077 | 0.972 | 0.927 | 0.952 | 0.080 | 0.977 |
| | dbCAN_2:HMMER | 0.923 | 0.947 | 0.076 | 0.971 | 0.936 | 0.950 | 0.060 | 0.963 | 0.940 | 0.953 | 0.039 | 0.965 |
| | dbCAN_2:DIAMOND | 0.945 | 0.970 | 0.080 | 0.996 | 0.949 | 0.967 | 0.082 | 0.985 | 0.937 | 0.964 | 0.085 | 0.991 |
| | dbCAN_2:Hotpep | 0.877 | 0.898 | 0.068 | 0.920 | 0.883 | 0.903 | 0.091 | 0.923 | 0.873 | 0.908 | 0.110 | 0.943 |
| | dbCAN_3 | 0.950 | 0.973 | 0.074 | 0.997 | 0.962 | 0.975 | 0.059 | 0.988 | 0.965 | 0.977 | 0.038 | 0.989 |
| | dbCAN_3:HMMER | 0.922 | 0.946 | 0.075 | 0.970 | 0.937 | 0.950 | 0.060 | 0.964 | 0.942 | 0.954 | 0.039 | 0.967 |
| | dbCAN_3:DIAMOND | 0.960 | 0.983 | 0.072 | 1.006 | 0.972 | 0.985 | 0.057 | 0.997 | 0.975 | 0.987 | 0.036 | 0.998 |
| | dbCAN_3:eCAMI | 0.932 | 0.952 | 0.064 | 0.973 | 0.923 | 0.941 | 0.082 | 0.960 | 0.899 | 0.930 | 0.097 | 0.961 |
| | dbCAN_4 | 0.953 | 0.976 | 0.073 | 1.000 | 0.965 | 0.978 | 0.057 | 0.991 | 0.969 | 0.980 | 0.035 | 0.991 |
| | dbCAN_4:HMMER | 0.922 | 0.947 | 0.075 | 0.971 | 0.937 | 0.950 | 0.060 | 0.964 | 0.942 | 0.954 | 0.039 | 0.966 |
| | dbCAN_4:DIAMOND | 0.956 | 0.980 | 0.074 | 1.003 | 0.971 | 0.984 | 0.058 | 0.997 | 0.977 | 0.988 | 0.035 | 0.999 |
| | dbCAN_4:dbCAN-sub | 0.956 | 0.977 | 0.065 | 0.998 | 0.966 | 0.977 | 0.052 | 0.989 | 0.966 | 0.978 | 0.035 | 0.989 |
| | CUPP | 0.945 | 0.960 | 0.049 | 0.976 | 0.933 | 0.946 | 0.058 | 0.959 | 0.912 | 0.932 | 0.063 | 0.953 |
| PL | dbCAN_2 | 0.993 | 0.996 | 0.008 | 0.999 | 0.994 | 0.996 | 0.007 | 0.998 | 0.993 | 0.996 | 0.006 | 0.999 |
| | dbCAN_2:HMMER | 0.994 | 0.996 | 0.007 | 0.999 | 0.994 | 0.996 | 0.006 | 0.998 | 0.994 | 0.997 | 0.005 | 1.000 |
| | dbCAN_2:DIAMOND | 0.993 | 0.996 | 0.008 | 0.999 | 0.993 | 0.995 | 0.007 | 0.998 | 0.992 | 0.995 | 0.007 | 0.998 |
| | dbCAN_2:Hotpep | 0.986 | 0.992 | 0.016 | 0.998 | 0.989 | 0.993 | 0.013 | 0.997 | 0.992 | 0.995 | 0.007 | 0.998 |
| | dbCAN_3 | 0.998 | 0.999 | 0.003 | 1.000 | 0.998 | 0.999 | 0.003 | 1.000 | 0.997 | 0.999 | 0.003 | 1.000 |
| | dbCAN_3:HMMER | 0.997 | 0.998 | 0.004 | 1.000 | 0.998 | 0.999 | 0.003 | 1.000 | 0.997 | 0.999 | 0.003 | 1.000 |
| | dbCAN_3:DIAMOND | 0.998 | 0.999 | 0.003 | 1.000 | 0.997 | 0.999 | 0.004 | 1.000 | 0.995 | 0.998 | 0.006 | 1.001 |
| | dbCAN_3:eCAMI | 0.985 | 0.991 | 0.015 | 0.997 | 0.988 | 0.992 | 0.013 | 0.996 | 0.990 | 0.994 | 0.006 | 0.997 |
| | dbCAN_4 | 0.997 | 0.998 | 0.004 | 1.000 | 0.998 | 0.999 | 0.003 | 1.000 | 0.997 | 0.999 | 0.003 | 1.000 |
| | dbCAN_4:HMMER | 0.997 | 0.998 | 0.004 | 1.000 | 0.998 | 0.999 | 0.003 | 1.000 | 0.997 | 0.999 | 0.003 | 1.000 |
| | dbCAN_4:DIAMOND | 0.998 | 0.999 | 0.003 | 1.000 | 0.998 | 0.999 | 0.004 | 1.000 | 0.996 | 0.998 | 0.005 | 1.001 |
| | dbCAN_4:dbCAN-sub | 0.997 | 0.999 | 0.003 | 1.000 | 0.998 | 0.999 | 0.003 | 1.000 | 0.997 | 0.999 | 0.003 | 1.000 |
| | CUPP | 0.989 | 0.993 | 0.011 | 0.997 | 0.991 | 0.994 | 0.010 | 0.997 | 0.993 | 0.996 | 0.006 | 0.999 |
| CE | dbCAN_2 | 0.976 | 0.985 | 0.028 | 0.994 | 0.984 | 0.989 | 0.023 | 0.994 | 0.990 | 0.994 | 0.012 | 0.998 |
| | dbCAN_2:HMMER | 0.976 | 0.984 | 0.024 | 0.992 | 0.986 | 0.990 | 0.019 | 0.994 | 0.995 | 0.997 | 0.006 | 0.999 |
| | dbCAN_2:DIAMOND | 0.973 | 0.982 | 0.028 | 0.991 | 0.981 | 0.986 | 0.024 | 0.992 | 0.984 | 0.990 | 0.018 | 0.997 |
| | dbCAN_2:Hotpep | 0.969 | 0.979 | 0.031 | 0.989 | 0.976 | 0.982 | 0.025 | 0.988 | 0.981 | 0.986 | 0.016 | 0.991 |
| | dbCAN_3 | 0.975 | 0.984 | 0.028 | 0.993 | 0.985 | 0.990 | 0.022 | 0.995 | 0.992 | 0.996 | 0.010 | 0.999 |
| | dbCAN_3:HMMER | 0.978 | 0.985 | 0.022 | 0.992 | 0.987 | 0.991 | 0.018 | 0.995 | 0.996 | 0.998 | 0.005 | 0.999 |
| | dbCAN_3:DIAMOND | 0.978 | 0.986 | 0.027 | 0.995 | 0.984 | 0.989 | 0.023 | 0.994 | 0.986 | 0.992 | 0.018 | 0.998 |
| | dbCAN_3:eCAMI | 0.968 | 0.978 | 0.030 | 0.987 | 0.976 | 0.982 | 0.025 | 0.987 | 0.981 | 0.986 | 0.016 | 0.992 |
| | dbCAN_4 | 0.983 | 0.990 | 0.023 | 0.998 | 0.990 | 0.994 | 0.017 | 0.998 | 0.997 | 0.999 | 0.004 | 1.000 |
| | dbCAN_4:HMMER | 0.981 | 0.988 | 0.023 | 0.995 | 0.989 | 0.993 | 0.017 | 0.997 | 0.997 | 0.998 | 0.005 | 1.000 |
| | dbCAN_4:DIAMOND | 0.980 | 0.989 | 0.026 | 0.997 | 0.987 | 0.992 | 0.022 | 0.997 | 0.991 | 0.996 | 0.015 | 1.001 |
| | dbCAN_4:dbCAN-sub | 0.983 | 0.990 | 0.023 | 0.997 | 0.990 | 0.994 | 0.017 | 0.998 | 0.996 | 0.998 | 0.006 | 1.000 |
| | CUPP | 0.978 | 0.985 | 0.024 | 0.993 | 0.986 | 0.990 | 0.019 | 0.994 | 0.992 | 0.995 | 0.009 | 0.998 |
| AA | dbCAN_2 | 1.000 | 1.000 | 0.000 | 1.000 | 0.980 | 0.986 | 0.020 | 0.991 | 0.974 | 0.981 | 0.022 | 0.988 |
| | dbCAN_2:HMMER | 1.000 | 1.000 | 0.000 | 1.000 | 0.980 | 0.986 | 0.020 | 0.991 | 0.974 | 0.981 | 0.021 | 0.988 |
| | dbCAN_2:DIAMOND | 1.000 | 1.000 | 0.000 | 1.000 | 0.978 | 0.984 | 0.021 | 0.990 | 0.972 | 0.979 | 0.022 | 0.986 |
| | dbCAN_2:Hotpep | 0.997 | 0.999 | 0.003 | 1.001 | 0.976 | 0.983 | 0.023 | 0.989 | 0.970 | 0.978 | 0.024 | 0.986 |
| | dbCAN_3 | 1.000 | 1.000 | 0.000 | 1.000 | 0.985 | 0.990 | 0.019 | 0.996 | 0.980 | 0.987 | 0.021 | 0.994 |
| | dbCAN_3:HMMER | 1.000 | 1.000 | 0.000 | 1.000 | 0.984 | 0.990 | 0.020 | 0.996 | 0.980 | 0.987 | 0.021 | 0.994 |
| | dbCAN_3:DIAMOND | 1.000 | 1.000 | 0.000 | 1.000 | 0.984 | 0.990 | 0.020 | 0.995 | 0.979 | 0.987 | 0.022 | 0.994 |
| | dbCAN_3:eCAMI | 0.997 | 0.999 | 0.003 | 1.001 | 0.972 | 0.979 | 0.023 | 0.986 | 0.965 | 0.973 | 0.023 | 0.981 |
| | dbCAN_4 | 1.000 | 1.000 | 0.000 | 1.000 | 0.985 | 0.991 | 0.019 | 0.996 | 0.981 | 0.988 | 0.021 | 0.995 |
| | dbCAN_4:HMMER | 1.000 | 1.000 | 0.000 | 1.000 | 0.984 | 0.990 | 0.020 | 0.996 | 0.980 | 0.987 | 0.021 | 0.994 |
| | dbCAN_4:DIAMOND | 0.997 | 0.999 | 0.003 | 1.001 | 0.984 | 0.990 | 0.020 | 0.995 | 0.980 | 0.987 | 0.022 | 0.994 |
| | dbCAN_4:dbCAN-sub | 1.000 | 1.000 | 0.000 | 1.000 | 0.985 | 0.990 | 0.019 | 0.996 | 0.980 | 0.987 | 0.021 | 0.994 |
| | CUPP | 1.000 | 1.000 | 0.000 | 1.000 | 0.978 | 0.984 | 0.022 | 0.990 | 0.972 | 0.979 | 0.022 | 0.987 |
| CBM | dbCAN_2 | 0.959 | 0.967 | 0.027 | 0.976 | 0.960 | 0.966 | 0.026 | 0.971 | 0.956 | 0.964 | 0.025 | 0.972 |
| | dbCAN_2:HMMER | 0.929 | 0.940 | 0.036 | 0.951 | 0.931 | 0.938 | 0.030 | 0.944 | 0.928 | 0.936 | 0.022 | 0.943 |
| | dbCAN_2:DIAMOND | 0.964 | 0.972 | 0.024 | 0.980 | 0.966 | 0.972 | 0.025 | 0.978 | 0.964 | 0.972 | 0.027 | 0.981 |
| | dbCAN_2:Hotpep | 0.853 | 0.873 | 0.063 | 0.894 | 0.870 | 0.882 | 0.055 | 0.895 | 0.877 | 0.891 | 0.045 | 0.905 |
| | dbCAN_3 | 0.963 | 0.971 | 0.024 | 0.978 | 0.967 | 0.972 | 0.023 | 0.977 | 0.966 | 0.973 | 0.022 | 0.980 |
| | dbCAN_3:HMMER | 0.928 | 0.940 | 0.036 | 0.952 | 0.931 | 0.938 | 0.030 | 0.945 | 0.929 | 0.936 | 0.022 | 0.943 |
| | dbCAN_3:DIAMOND | 0.970 | 0.976 | 0.018 | 0.982 | 0.977 | 0.981 | 0.019 | 0.986 | 0.981 | 0.987 | 0.018 | 0.993 |
| | dbCAN_3:eCAMI | 0.898 | 0.917 | 0.060 | 0.936 | 0.914 | 0.927 | 0.056 | 0.939 | 0.920 | 0.936 | 0.051 | 0.953 |
| | dbCAN_4 | 0.981 | 0.985 | 0.013 | 0.990 | 0.972 | 0.976 | 0.021 | 0.981 | 0.960 | 0.967 | 0.023 | 0.975 |
| | dbCAN_4:HMMER | 0.938 | 0.949 | 0.036 | 0.961 | 0.937 | 0.944 | 0.030 | 0.950 | 0.931 | 0.938 | 0.021 | 0.944 |
| | dbCAN_4:DIAMOND | 0.970 | 0.975 | 0.018 | 0.981 | 0.979 | 0.983 | 0.018 | 0.987 | 0.986 | 0.991 | 0.015 | 0.995 |
| | dbCAN_4:dbCAN-sub | 0.985 | 0.989 | 0.013 | 0.993 | 0.971 | 0.976 | 0.024 | 0.982 | 0.955 | 0.963 | 0.026 | 0.971 |
| | CUPP | 0.825 | 0.861 | 0.113 | 0.897 | 0.866 | 0.885 | 0.085 | 0.904 | 0.900 | 0.909 | 0.029 | 0.918 |

## SI figure 32: The mean accuracy and 95% confidence interval of the CAZy class classification across all CAZy classes (overleaf)

Accuracy was calculated for each test set across all CAZy class. The mean accuracy was then calculated across all test sets and 95% confidence interval (CI) was across all CAZy class, and is plotted in SI figure 32.

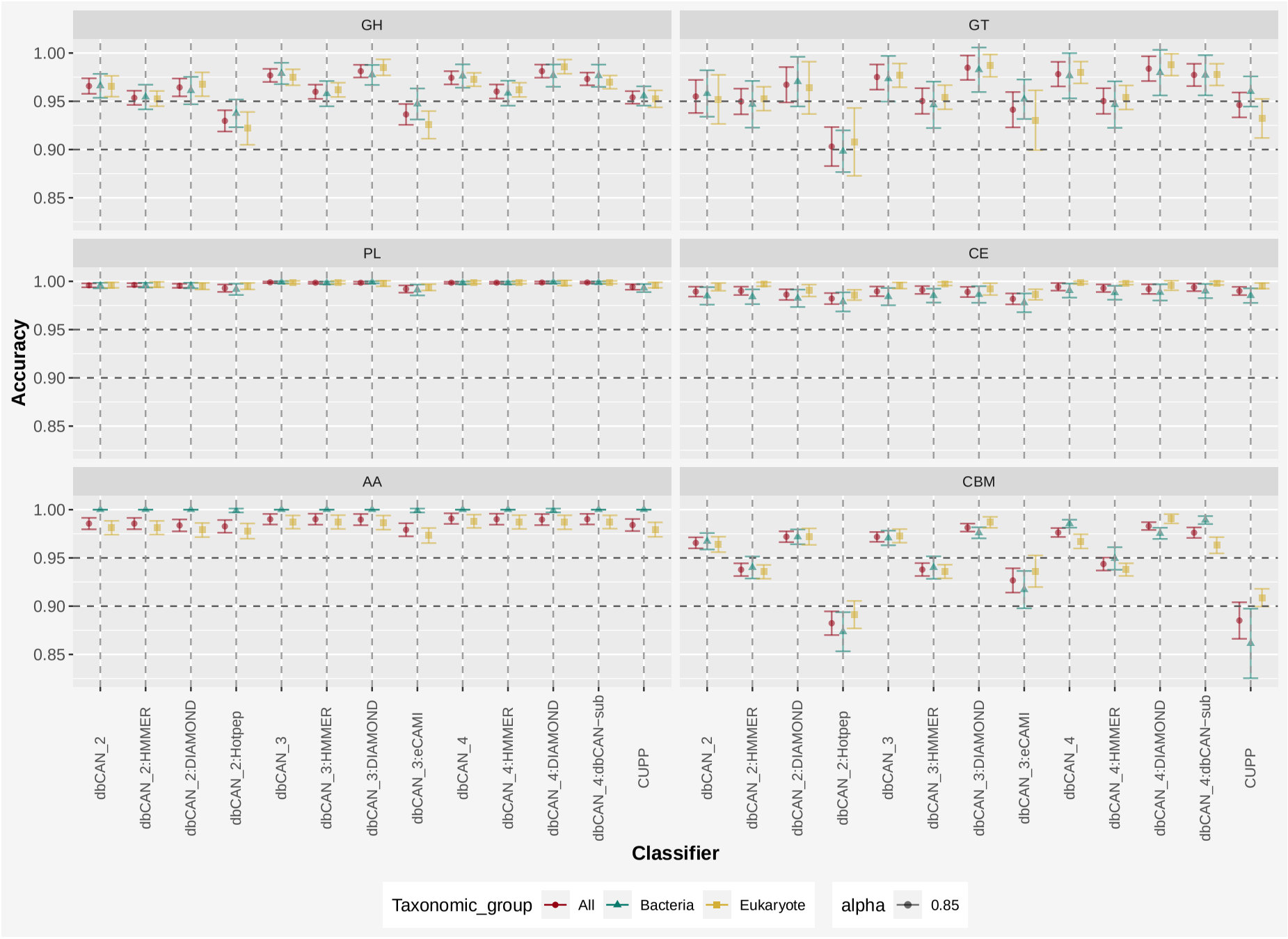



Figure 32: The mean accuracy and 95% confidence interval (CI) of binary CAZy class classification across all CAZy classes and across all test sets. Where a tool is incorporated into a CAZyme classifier, the tool name is prefixed with the respective CAZyme classifier. Per classifier (from left to right) all test sets are pooled together (shaded red); all bacterial test sets are pooled (shaded green); and all test sets from eukaryotic genomes are pooled together (shaded yellow).

**SI figure 33: The accuracy per test set of the CAZy class classification across all CAZy classes (overleaf)**

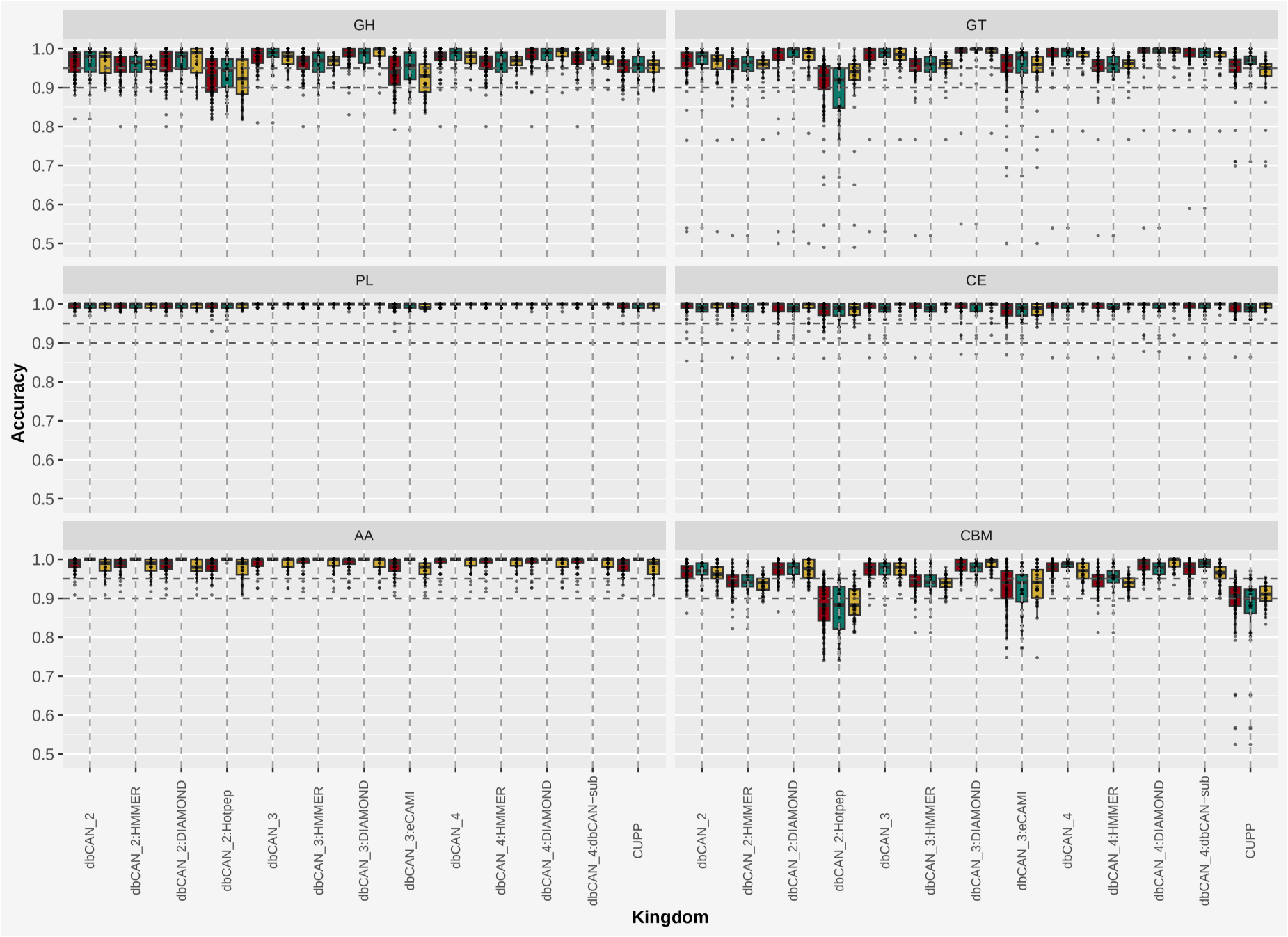



Figure 33: One-dimensional scatter plot overlaying a box and whisker plot, where each point represents the accuracy of the binary CAZy class classification across all CAZy classes and per test set. Where a tool is incorporated into a CAZyme classifier, the tool name is prefixed with the respective CAZyme classifier. Per classifier (from left to right) all test sets are pooled together (shaded red); all bacterial test sets are pooled (shaded green); and all test sets from eukaryotic genomes are pooled together (shaded yellow).

# 9 Testing for statically significant differences between taxonomic kingdoms for CAZy class classification

For each CAZy class a two-way ANOVA was performed to test for statistically significant differences between the tools and taxonomic kingdoms for each static (specificity, sensitivity, etc.). The output of these two-way ANOVAs are presented here. Where a statistically significant difference was detected, a Tukey HSD test was performed to determine between which groups the means were significantly different. Owing to their size, these results are available in the online repository.

## 9.1 Specificity

**SI Table 57: Output of a two-way ANOVA testing for statistically significant differences between specificity of tools and the taxonomic kingdom when classify GH CAZy class domains**

| Term | Degrees of Freedom | Sum of Squares | Mean Square | F-Statistic | P-value |
|---|---|---|---|---|---|
| Tax_group | 2 | 9.83E-05 | 4.92E-05 | 0.076147 | 0.926683 |
| Prediction_tool | 12 | 0.031897 | 0.002658 | 4.117436 | 2.14E-06 |
| Tax_group:Prediction_tool | 24 | 0.009063 | 0.000378 | 0.584928 | 0.945157 |
| Residuals | 2041 | 1.317583 | 0.000646 | NA | NA |

**SI Table 58: Output of a two-way ANOVA testing for statistically significant differences between specificity of tools and the taxonomic kingdom when classify GT CAZy class domains**

| Term | Degrees of Freedom | Sum of Squares | Mean Square | F-Statistic | P-value |
|---|---|---|---|---|---|
| Tax_group | 2 | 0.061499 | 0.03075 | 13.91087 | 9.99E-07 |
| Prediction_tool | 12 | 0.002773 | 0.000231 | 0.104555 | 0.99995 |
| Tax_group:Prediction_tool | 24 | 0.001765 | 7.35E-05 | 0.033273 | 1 |
| Residuals | 2041 | 4.511593 | 0.00221 | NA | NA |

**SI Table 59: Output of a two-way ANOVA testing for statistically significant differences between specificity of tools and the taxonomic kingdom when classify PL CAZy class domains**

| Term | Degrees of Freedom | Sum of Squares | Mean Square | F-Statistic | P-value |
|---|---|---|---|---|---|
| Tax_group | 2 | 0.061499 | 0.03075 | 13.91087 | 9.99E-07 |
| Prediction_tool | 12 | 0.002773 | 0.000231 | 0.104555 | 0.99995 |
| Tax_group:Prediction_tool | 24 | 0.001765 | 7.35E-05 | 0.033273 | 1 |
| Residuals | 2041 | 4.511593 | 0.00221 | NA | NA |

**SI Table 59: Output of a two-way ANOVA testing for statistically significant differences between specificity of tools and the taxonomic kingdom when classify CE CAZy class domains**

| Term | Degrees of Freedom | Sum of Squares | Mean Square | F-Statistic | P-value |
|---|---|---|---|---|---|
| Tax_group | 2 | 0.020416 | 0.010208 | 24.99738 | 1.91E-11 |
| Prediction_tool | 12 | 0.004011 | 0.000334 | 0.818495 | 0.631539 |
| Tax_group:Prediction_tool | 24 | 0.002107 | 8.78E-05 | 0.215014 | 0.999982 |
| Residuals | 1937 | 0.791003 | 0.000408 | NA | NA |

**SI Table 60: Output of a two-way ANOVA testing for statistically significant differences between specificity of tools and the taxonomic kingdom when classify AA CAZy class domains**

| Term | Degrees of Freedom | Sum of Squares | Mean Square | F-Statistic | P-value |
|---|---|---|---|---|---|
| Tax_group | 2 | 0.014462 | 0.007231 | 19.70271 | 3.80E-09 |
| Prediction_tool | 12 | 0.000688 | 5.73E-05 | 0.156142 | 0.999569 |
| Tax_group:Prediction_tool | 24 | 0.000102 | 4.26E-06 | 0.011605 | 1 |
| Residuals | 1209 | 0.443695 | 0.000367 | NA | NA |

**SI Table 61: Output of a two-way ANOVA testing for statistically significant differences between specificity of tools and the taxonomic kingdom when classify CBM CAZy class domains**

| Term | Degrees of Freedom | Sum of Squares | Mean Square | F-Statistic | P-value |
|---|---|---|---|---|---|
| Tax_group | 2 | 0.001859 | 0.00093 | 1.806734 | 0.164452 |
| Prediction_tool | 12 | 1.515693 | 0.126308 | 245.5026 | 0 |
| Tax_group:Prediction_tool | 24 | 0.018835 | 0.000785 | 1.525367 | 0.049277 |
| Residuals | 2041 | 1.050067 | 0.000514 | NA | NA |

## 9.2 Precision

**SI Table 62: Output of a two-way ANOVA testing for statistically significant differences between precision of tools and the taxonomic kingdom when classify GH CAZy class domains**

| Term | Degrees of Freedom | Sum of Squares | Mean Square | F-Statistic | P-value |
|---|---|---|---|---|---|
| Tax_group | 2 | 0.002808 | 0.001404 | 1.088348 | 0.336968 |
| Prediction_tool | 12 | 0.028378 | 0.002365 | 1.832953 | 0.038297 |
| Tax_group:Prediction_tool | 24 | 0.008193 | 0.000341 | 0.2646 | 0.999875 |
| Residuals | 2041 | 2.633262 | 0.00129 | NA | NA |

**SI Table 63: Output of a two-way ANOVA testing for statistically significant differences between precision of tools and the taxonomic kingdom when classify GT CAZy class domains**

| Term | Degrees of Freedom | Sum of Squares | Mean Square | F-Statistic | P-value |
|---|---|---|---|---|---|
| Tax_group | 2 | 0.113651 | 0.056825 | 16.52065 | 7.63E-08 |
| Prediction_tool | 12 | 0.008562 | 0.000714 | 0.207438 | 0.998184 |
| Tax_group:Prediction_tool | 24 | 0.003231 | 0.000135 | 0.039142 | 1 |
| Residuals | 2041 | 7.02034 | 0.00344 | NA | NA |

**SI Table 64: Output of a two-way ANOVA testing for statistically significant differences between precision of tools and the taxonomic kingdom when classify PL CAZy class domains**

| Term | Degrees of Freedom | Sum of Squares | Mean Square | F-Statistic | P-value |
|---|---|---|---|---|---|
| Tax_group | 2 | 0.438973 | 0.219487 | 8.934397 | 0.000141 |
| Prediction_tool | 12 | 0.503894 | 0.041991 | 1.709288 | 0.05952 |
| Tax_group:Prediction_tool | 24 | 0.112691 | 0.004695 | 0.191133 | 0.999994 |
| Residuals | 1185 | 29.11126 | 0.024566 | NA | NA |

**SI Table 65: Output of a two-way ANOVA testing for statistically significant differences between precision of tools and the taxonomic kingdom when classify CE CAZy class domains**

| Term | Degrees of Freedom | Sum of Squares | Mean Square | F-Statistic | P-value |
|---|---|---|---|---|---|
| Tax_group | 2 | 0.973304 | 0.486652 | 23.02689 | 1.31E-10 |
| Prediction_tool | 12 | 0.56175 | 0.046812 | 2.215025 | 0.009211 |
| Tax_group:Prediction_tool | 24 | 0.182517 | 0.007605 | 0.359839 | 0.998227 |
| Residuals | 1937 | 40.93669 | 0.021134 | NA | NA |

**SI Table 66: Output of a two-way ANOVA testing for statistically significant differences between precision of tools and the taxonomic kingdom when classify AA CAZy class domains**

| Term | Degrees of Freedom | Sum of Squares | Mean Square | F-Statistic | P-value |
|---|---|---|---|---|---|
| Tax_group | 2 | 0.993031 | 0.496516 | 15.37944 | 2.54E-07 |
| Prediction_tool | 12 | 0.43671 | 0.036392 | 1.127248 | 0.333264 |
| Tax_group:Prediction_tool | 24 | 0.067413 | 0.002809 | 0.087004 | 1 |
| Residuals | 1209 | 39.0318 | 0.032284 | NA | NA |

**SI Table 67: Output of a two-way ANOVA testing for statistically significant differences between precision of tools and the taxonomic kingdom when classify CBM CAZy class domains**

| Term | Degrees of Freedom | Sum of Squares | Mean Square | F-Statistic | P-value |
|---|---|---|---|---|---|
| Tax_group | 2 | 0.264912 | 0.132456 | 5.098048 | 0.006187 |
| Prediction_tool | 12 | 150.651 | 12.55425 | 483.1951 | 0 |
| Tax_group:Prediction_tool | 24 | 0.421297 | 0.017554 | 0.675629 | 0.879297 |
| Residuals | 2041 | 53.02873 | 0.025982 | NA | NA |

## 9.3 Sensitivity

**SI Table 62: Output of a two-way ANOVA testing for statistically significant differences between sensitivity of tools and the taxonomic kingdom when classify GH CAZy class domains**

| Term | Degrees of Freedom | Sum of Squares | Mean Square | F-Statistic | P-value |
|---|---|---|---|---|---|
| Tax_group | 2 | 0.024777 | 0.012389 | 1.633256 | 0.195548 |
| Prediction_tool | 12 | 2.223371 | 0.185281 | 24.42671 | 1.01E-51 |
| Tax_group:Prediction_tool | 24 | 0.223976 | 0.009332 | 1.23034 | 0.202792 |
| Residuals | 2041 | 15.48135 | 0.007585 | NA | NA |

**SI Table 63: Output of a two-way ANOVA testing for statistically significant differences between sensitivity of tools and the taxonomic kingdom when classify GT CAZy class domains**

| Term | Degrees of Freedom | Sum of Squares | Mean Square | F-Statistic | P-value |
|---|---|---|---|---|---|
| Tax_group | 2 | 0.113119 | 0.056559 | 3.920567 | 0.019979 |
| Prediction_tool | 12 | 9.463921 | 0.78866 | 54.66821 | 2.43E-114 |
| Tax_group:Prediction_tool | 24 | 0.236102 | 0.009838 | 0.68192 | 0.873545 |
| Residuals | 2041 | 29.44408 | 0.014426 | NA | NA |

**SI Table 64: Output of a two-way ANOVA testing for statistically significant differences between sensitivity of tools and the taxonomic kingdom when classify PL CAZy class domains**

| Term | Degrees of Freedom | Sum of Squares | Mean Square | F-Statistic | P-value |
|---|---|---|---|---|---|
| Tax_group | 2 | 0.072826 | 0.036413 | 1.087647 | 0.337345 |
| Prediction_tool | 12 | 5.505987 | 0.458832 | 13.70522 | 6.45E-27 |
| Tax_group:Prediction_tool | 24 | 0.537249 | 0.022385 | 0.668646 | 0.884898 |
| Residuals | 1185 | 39.67221 | 0.033479 | NA | NA |

**SI Table 65: Output of a two-way ANOVA testing for statistically significant differences between sensitivity of tools and the taxonomic kingdom when classify CE CAZy class domains**

| Term | Degrees of Freedom | Sum of Squares | Mean Square | F-Statistic | P-value |
|---|---|---|---|---|---|
| Tax_group | 2 | 5.93E-05 | 2.96E-05 | 0.001238 | 0.998763 |
| Prediction_tool | 12 | 5.085574 | 0.423798 | 17.70147 | 1.22E-36 |
| Tax_group:Prediction_tool | 24 | 1.219281 | 0.050803 | 2.12199 | 0.001203 |
| Residuals | 1937 | 46.37447 | 0.023941 | NA | NA |

**SI Table 66: Output of a two-way ANOVA testing for statistically significant differences between sensitivity of tools and the taxonomic kingdom when classify AA CAZy class domains**

| Term | Degrees of Freedom | Sum of Squares | Mean Square | F-Statistic | P-value |
|---|---|---|---|---|---|
| Tax_group | 2 | 0.376549 | 0.188274 | 13.07234 | 2.42E-06 |
| Prediction_tool | 12 | 4.89413 | 0.407844 | 28.31759 | 3.39E-57 |
| Tax_group:Prediction_tool | 24 | 0.409773 | 0.017074 | 1.18548 | 0.244467 |
| Residuals | 1209 | 17.41263 | 0.014403 | NA | NA |

**SI Table 67: Output of a two-way ANOVA testing for statistically significant differences between sensitivity of tools and the taxonomic kingdom when classify CBM CAZy class domains**

| Term | Degrees of Freedom | Sum of Squares | Mean Square | F-Statistic | P-value |
|---|---|---|---|---|---|
| Tax_group | 2 | 0.941706 | 0.470853 | 13.19133 | 2.03E-06 |
| Prediction_tool | 12 | 122.1305 | 10.17754 | 285.1321 | 0 |
| Tax_group:Prediction_tool | 24 | 5.042356 | 0.210098 | 5.886072 | 7.21E-18 |
| Residuals | 2041 | 72.8517 | 0.035694 | NA | NA |

## 9.4 F1-score

**SI Table 74: Output of a two-way ANOVA testing for statistically significant differences between F1-score of tools and the taxonomic kingdom when classify GH CAZy class domains**

| Term | Degrees of Freedom | Sum of Squares | Mean Square | F-Statistic | P-value |
|---|---|---|---|---|---|
| Tax_group | 2 | 0.006578 | 0.003289 | 0.887998 | 0.411637 |
| Prediction_tool | 12 | 0.714402 | 0.059534 | 16.07316 | 4.52E-33 |
| Tax_group:Prediction_tool | 24 | 0.060311 | 0.002513 | 0.678459 | 0.876728 |
| Residuals | 2041 | 7.559675 | 0.003704 | NA | NA |

**SI Table 75: Output of a two-way ANOVA testing for statistically significant differences between F1-score of tools and the taxonomic kingdom when classify GT CAZy class domains**

| Term | Degrees of Freedom | Sum of Squares | Mean Square | F-Statistic | P-value |
|---|---|---|---|---|---|
| Tax_group | 2 | 0.00407 | 0.002035 | 0.243723 | 0.783728 |
| Prediction_tool | 12 | 3.45841 | 0.288201 | 34.51828 | 1.68E-73 |
| Tax_group:Prediction_tool | 24 | 0.114905 | 0.004788 | 0.57343 | 0.951254 |
| Residuals | 2041 | 17.04076 | 0.008349 | NA | NA |

**SI Table 76: Output of a two-way ANOVA testing for statistically significant differences between F1-score of tools and the taxonomic kingdom when classify PL CAZy class domains**

| Term | Degrees of Freedom | Sum of Squares | Mean Square | F-Statistic | P-value |
|---|---|---|---|---|---|
| Tax_group | 2 | 0.239401 | 0.119701 | 4.447763 | 0.011901 |
| Prediction_tool | 12 | 2.695021 | 0.224585 | 8.344987 | 2.65E-15 |
| Tax_group:Prediction_tool | 24 | 0.262499 | 0.010937 | 0.406406 | 0.995309 |
| Residuals | 1185 | 31.8914 | 0.026913 | NA | NA |

**SI Table 77: Output of a two-way ANOVA testing for statistically significant differences between F1-score of tools and the taxonomic kingdom when classify CE CAZy class domains**

| Term | Degrees of Freedom | Sum of Squares | Mean Square | F-Statistic | P-value |
|---|---|---|---|---|---|
| Tax_group | 2 | 0.322581 | 0.161291 | 8.207119 | 0.000282 |
| Prediction_tool | 12 | 3.141802 | 0.261817 | 13.3223 | 8.71E-27 |
| Tax_group:Prediction_tool | 24 | 0.422647 | 0.01761 | 0.896083 | 0.608565 |
| Residuals | 1937 | 38.06695 | 0.019653 | NA | NA |

**SI Table 78: Output of a two-way ANOVA testing for statistically significant differences between F1-score of tools and the taxonomic kingdom when classify AA CAZy class domains**

| Term | Degrees of Freedom | Sum of Squares | Mean Square | F-Statistic | P-value |
|---|---|---|---|---|---|
| Tax_group | 2 | 0.922797 | 0.461398 | 23.28009 | 1.20E-10 |
| Prediction_tool | 12 | 2.075687 | 0.172974 | 8.727487 | 3.79E-16 |
| Tax_group:Prediction_tool | 24 | 0.12771 | 0.005321 | 0.268487 | 0.999854 |
| Residuals | 1209 | 23.9617 | 0.019819 | NA | NA |

**SI Table 79: Output of a two-way ANOVA testing for statistically significant differences between F1-score of tools and the taxonomic kingdom when classify CBM CAZy class domains**

| Term | Degrees of Freedom | Sum of Squares | Mean Square | F-Statistic | P-value |
|---|---|---|---|---|---|
| Tax_group | 2 | 1.00197 | 0.500985 | 18.69021 | 9.04E-09 |
| Prediction_tool | 12 | 119.4761 | 9.956342 | 371.4404 | 0 |
| Tax_group:Prediction_tool | 24 | 2.940975 | 0.122541 | 4.571613 | 1.57E-12 |
| Residuals | 2041 | 54.70836 | 0.026805 | NA | NA |

## 9.5 Accuracy

**SI Table 80: Output of a two-way ANOVA testing for statistically significant differences between accuracy of tools and the taxonomic kingdom when classify GH CAZy class domains**

| Term | Degrees of Freedom | Sum of Squares | Mean Square | F-Statistic | P-value |
|---|---|---|---|---|---|
| Tax_group | 2 | 0.000995 | 0.000497 | 0.392998 | 0.675081 |
| Prediction_tool | 12 | 0.495514 | 0.041293 | 32.62116 | 1.79E-69 |
| Tax_group:Prediction_tool | 24 | 0.019307 | 0.000804 | 0.635507 | 0.912384 |
| Residuals | 2041 | 2.583556 | 0.001266 | NA | NA |

**SI Table 81: Output of a two-way ANOVA testing for statistically significant differences between accuracy of tools and the taxonomic kingdom when classify GT CAZy class domains**

| Term | Degrees of Freedom | Sum of Squares | Mean Square | F-Statistic | P-value |
|---|---|---|---|---|---|
| Tax_group | 2 | 0.000201 | 0.0001 | 0.022561 | 0.977692 |
| Prediction_tool | 12 | 0.986214 | 0.082184 | 18.48451 | 1.67E-38 |
| Tax_group:Prediction_tool | 24 | 0.033444 | 0.001394 | 0.313422 | 0.999442 |
| Residuals | 2041 | 9.074547 | 0.004446 | NA | NA |

**SI Table 82: Output of a two-way ANOVA testing for statistically significant differences between accuracy of tools and the taxonomic kingdom when classify PL CAZy class domains**

| Term | Degrees of Freedom | Sum of Squares | Mean Square | F-Statistic | P-value |
|---|---|---|---|---|---|
| Tax_group | 2 | 8.05E-05 | 4.02E-05 | 0.787336 | 0.455293 |
| Prediction_tool | 12 | 0.006872 | 0.000573 | 11.20334 | 1.64E-21 |
| Tax_group:Prediction_tool | 24 | 0.000279 | 1.16E-05 | 0.227095 | 0.999969 |
| Residuals | 1185 | 0.060569 | 5.11E-05 | NA | NA |

**SI Table 83: Output of a two-way ANOVA testing for statistically significant differences between accuracy of tools and the taxonomic kingdom when classify CE CAZy class domains**

| Term | Degrees of Freedom | Sum of Squares | Mean Square | F-Statistic | P-value |
|---|---|---|---|---|---|
| Tax_group | 2 | 0.020684 | 0.010342 | 23.83656 | 5.93E-11 |
| Prediction_tool | 12 | 0.028397 | 0.002366 | 5.454192 | 3.32E-09 |
| Tax_group:Prediction_tool | 24 | 0.001068 | 4.45E-05 | 0.102525 | 1 |
| Residuals | 1937 | 0.840419 | 0.000434 | NA | NA |

**SI Table 84: Output of a two-way ANOVA testing for statistically significant differences between accuracy of tools and the taxonomic kingdom when classify AA CAZy class domains**

| Term | Degrees of Freedom | Sum of Squares | Mean Square | F-Statistic | P-value |
|---|---|---|---|---|---|
| Tax_group | 2 | 0.030258 | 0.015129 | 38.37394 | 6.95E-17 |
| Prediction_tool | 12 | 0.016442 | 0.00137 | 3.475235 | 4.59E-05 |
| Tax_group:Prediction_tool | 24 | 0.002225 | 9.27E-05 | 0.235096 | 0.999957 |
| Residuals | 1209 | 0.476657 | 0.000394 | NA | NA |

**SI Table 85: Output of a two-way ANOVA testing for statistically significant differences between accuracy of tools and the taxonomic kingdom when classify CBM CAZy class domains**

| Term | Degrees of Freedom | Sum of Squares | Mean Square | F-Statistic | P-value |
|---|---|---|---|---|---|
| Tax_group | 2 | 0.003097 | 0.001548 | 1.047743 | 0.350917 |
| Prediction_tool | 12 | 2.289676 | 0.190806 | 129.1232 | 6.47E-240 |
| Tax_group:Prediction_tool | 24 | 0.08645 | 0.003602 | 2.437629 | 0.000122 |
| Residuals | 2041 | 3.016 | 0.001478 | NA | NA |

### 9.6 Statistically significant differences

**SI table 86: Accumulated outputs of Tukey HSD performed per CAZy class for each statistic, reporting only where the CAZy class, tool (classifier) and taxonomic kingdom of the test set are the same and a statistically significant difference between the kingdoms if found (p-value<0.05) (overleaf)**

| CAZy Class | Statistic | Kingdom of Classifier 1 | Classifier 1 | Kingdom of Classifier 2 | Classifier 2 | Mean Difference | Lower 95% Confidence Interval | Upper 95% Confidence Interval | Adjusted P-value |
|---|---|---|---|---|---|---|---|---|---|
| GH | Sensitivity | Eukaryote | dbCAN_3:eCAMI | Bacteria | dbCAN_3:eCAMI | -0.075657664 | -0.1512499 | -6.54E-05 | 0.049420554 |
| CE | Sensitivity | Eukaryote | dbCAN_2:Hotpep | Bacteria | dbCAN_2:Hotpep | -0.151640212 | -0.28962958 | -0.013650843 | 0.011772207 |
| AA | Sensitivity | Eukaryote | dbCAN_3:eCAMI | Bacteria | dbCAN_3:eCAMI | -0.16478274 | -0.32492801 | -0.00463747 | 0.033569016 |
| AA | Sensitivity | Eukaryote | dbCAN_2:DIAMOND | Bacteria | dbCAN_2:DIAMOND | -0.16953662 | -0.329681889 | -0.00939135 | 0.021861947 |
| CBM | Sensitivity | Eukaryote | dbCAN_4:sub | Bacteria | dbCAN_4:sub | -0.249609445 | -0.413590121 | -0.085628768 | 2.94E-06 |
| CBM | Sensitivity | Eukaryote | dbCAN_3:HMMER | Bacteria | dbCAN_3:HMMER | -0.191503965 | -0.355484642 | -0.027523289 | 0.003788629 |
| CBM | Sensitivity | Eukaryote | dbCAN_4 | Bacteria | dbCAN_4 | -0.205949178 | -0.369929854 | -0.041968502 | 0.000778628 |
| CBM | Sensitivity | Eukaryote | dbCAN_4:HMMER | Bacteria | dbCAN_4:HMMER | -0.269098268 | -0.433078944 | -0.105117591 | 1.72E-07 |
| CBM | Sensitivity | Eukaryote | dbCAN_2:HMMER | Bacteria | dbCAN_2:HMMER | -0.195420361 | -0.359401038 | -0.031439685 | 0.002502563 |
| CBM | Specificity | Eukaryote | dbCAN_3:eCAMI | Bacteria | dbCAN_3:eCAMI | 0.024190081 | 0.004503002 | 0.04387716 | 0.001288752 |
| CBM | F1-score | Eukaryote | dbCAN_4:sub | Bacteria | dbCAN_4:sub | -0.181550001 | -0.323651832 | -0.03944817 | 0.000516369 |
| CBM | F1-score | Eukaryote | dbCAN_3:HMMER | Bacteria | dbCAN_3:HMMER | -0.18254838 | -0.324650211 | -0.040446549 | 0.000451196 |
| CBM | F1-score | Eukaryote | dbCAN_4 | Bacteria | dbCAN_4 | -0.149806035 | -0.291907866 | -0.007704204 | 0.023264282 |
| CBM | F1-score | Eukaryote | dbCAN_4:HMMER | Bacteria | dbCAN_4:HMMER | -0.24180764 | -0.383909471 | -0.099705809 | 3.73E-08 |
| CBM | F1-score | Eukaryote | dbCAN_2:HMMER | Bacteria | dbCAN_2:HMMER | -0.186770683 | -0.328872514 | -0.044668852 | 0.000252643 |
| CBM | Accuracy | Eukaryote | CUPP | Bacteria | CUPP | 0.047459703 | 0.014094905 | 0.080824501 | 2.71E-05 |

# 10 Multi-label classification of CAZy classes

## SI Table 87: Rand Index and Adjusted Rand Index of CAZy class multi-label classification

CAZy and CAZyme classifiers classify protein sequences in a domain-wise manner, assigning a CAZy family annotation for each CAZyme domain. Therefore, a protein sequence can be assigned to multiple CAZy families, and thus can be assigned to multiple CAZy classes. The performance of this multi-label classification of CAZy classes is evaluated by calculating the Rand Index (RI), ranging from 1 to represent always correct (i.e. assigning all required CAZy class annotations correctly) to 0 (where all classifications are incorrect). Additionally, the Adjusted Rand Index (ARI) is calculated. The ARI is the RI adjusted for chance, and ranges from -1 (representing systematically incorrect classifications), through 0 (representing an accuracy of classification that is equivalent to randomly assigning CAZy class classifications), to 1 (representing all classifications are correct). The RI and ARI was calculated per test set. The mean RI and ARI was then calculated (as well as the standard deviation and 95% confidence interval (CI) of the mean) across all test sets, all test sets derived from bacterial genomes, and all tests generated from eukaryotic genomes.

| Rand Index | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Bacteria | | | | All | | | | Eukaryote | | | |
| Classifier | Lower CI | Mean | SD | Upper CI | Lower CI | Mean | SD | Upper CI | Lower CI | Mean | SD | Upper CI |
| dbCAN_2 | 0.977 | 0.978 | 0.082 | 0.980 | 0.976 | 0.977 | 0.086 | 0.978 | 0.973 | 0.975 | 0.088 | 0.977 |
| dbCAN_2:HMMER | 0.968 | 0.970 | 0.096 | 0.972 | 0.968 | 0.969 | 0.097 | 0.971 | 0.967 | 0.969 | 0.098 | 0.971 |
| dbCAN_2:DIAMOND | 0.978 | 0.980 | 0.080 | 0.982 | 0.978 | 0.979 | 0.082 | 0.980 | 0.976 | 0.978 | 0.084 | 0.980 |
| dbCAN_2:Hotpep | 0.944 | 0.947 | 0.125 | 0.949 | 0.944 | 0.946 | 0.125 | 0.948 | 0.944 | 0.946 | 0.125 | 0.949 |
| dbCAN_3 | 0.982 | 0.984 | 0.072 | 0.986 | 0.983 | 0.984 | 0.072 | 0.985 | 0.982 | 0.984 | 0.072 | 0.985 |
| dbCAN_3:HMMER | 0.969 | 0.971 | 0.094 | 0.973 | 0.970 | 0.971 | 0.094 | 0.973 | 0.970 | 0.972 | 0.093 | 0.974 |
| dbCAN_3:DIAMOND | 0.985 | 0.987 | 0.066 | 0.988 | 0.986 | 0.987 | 0.064 | 0.988 | 0.987 | 0.988 | 0.062 | 0.990 |
| dbCAN_3:eCAMI | 0.962 | 0.964 | 0.105 | 0.967 | 0.959 | 0.961 | 0.109 | 0.963 | 0.955 | 0.958 | 0.113 | 0.960 |
| dbCAN_4 | 0.986 | 0.988 | 0.063 | 0.989 | 0.985 | 0.986 | 0.068 | 0.987 | 0.982 | 0.983 | 0.073 | 0.985 |
| dbCAN_4:HMMER | 0.971 | 0.973 | 0.091 | 0.975 | 0.971 | 0.973 | 0.092 | 0.974 | 0.970 | 0.972 | 0.093 | 0.974 |
| dbCAN_4:DIAMOND | 0.985 | 0.986 | 0.067 | 0.988 | 0.987 | 0.988 | 0.063 | 0.989 | 0.989 | 0.990 | 0.058 | 0.991 |
| dbCAN_4:dbCAN-sub | 0.987 | 0.988 | 0.061 | 0.990 | 0.984 | 0.985 | 0.069 | 0.986 | 0.980 | 0.982 | 0.076 | 0.983 |
| CUPP | 0.957 | 0.959 | 0.110 | 0.962 | 0.958 | 0.960 | 0.109 | 0.961 | 0.958 | 0.960 | 0.109 | 0.963 |
| **Adjusted Rand Index** | | | | | | | | | | | | |
| | Bacteria | | | | All | | | | Eukaryote | | | |
| Classifier | Lower CI | Mean | SD | Upper CI | Lower CI | Mean | SD | Upper CI | Lower CI | Mean | SD | Upper CI |
| dbCAN_2 | 0.933 | 0.938 | 0.237 | 0.944 | 0.930 | 0.934 | 0.245 | 0.938 | 0.924 | 0.929 | 0.254 | 0.935 |
| dbCAN_2:HMMER | 0.911 | 0.917 | 0.269 | 0.923 | 0.911 | 0.916 | 0.270 | 0.920 | 0.909 | 0.914 | 0.271 | 0.920 |
| dbCAN_2:DIAMOND | 0.938 | 0.943 | 0.229 | 0.948 | 0.937 | 0.940 | 0.234 | 0.944 | 0.932 | 0.937 | 0.240 | 0.943 |
| dbCAN_2:Hotpep | 0.854 | 0.861 | 0.328 | 0.869 | 0.853 | 0.858 | 0.334 | 0.864 | 0.848 | 0.855 | 0.339 | 0.863 |
| dbCAN_3 | 0.950 | 0.955 | 0.204 | 0.959 | 0.951 | 0.954 | 0.206 | 0.957 | 0.949 | 0.954 | 0.208 | 0.958 |
| dbCAN_3:HMMER | 0.914 | 0.920 | 0.265 | 0.925 | 0.917 | 0.921 | 0.261 | 0.925 | 0.917 | 0.923 | 0.257 | 0.929 |
| dbCAN_3:DIAMOND | 0.959 | 0.963 | 0.187 | 0.967 | 0.962 | 0.965 | 0.183 | 0.967 | 0.963 | 0.966 | 0.179 | 0.970 |
| dbCAN_3:eCAMI | 0.900 | 0.907 | 0.278 | 0.913 | 0.890 | 0.894 | 0.298 | 0.899 | 0.875 | 0.882 | 0.316 | 0.889 |
| dbCAN_4 | 0.961 | 0.965 | 0.183 | 0.969 | 0.957 | 0.960 | 0.193 | 0.963 | 0.950 | 0.955 | 0.202 | 0.959 |
| dbCAN_4:HMMER | 0.919 | 0.925 | 0.259 | 0.930 | 0.920 | 0.924 | 0.258 | 0.928 | 0.918 | 0.924 | 0.256 | 0.930 |
| dbCAN_4:DIAMOND | 0.957 | 0.961 | 0.190 | 0.965 | 0.963 | 0.966 | 0.180 | 0.969 | 0.967 | 0.970 | 0.168 | 0.974 |
| dbCAN_4:dbCAN-sub | 0.963 | 0.967 | 0.177 | 0.971 | 0.955 | 0.958 | 0.195 | 0.961 | 0.945 | 0.950 | 0.212 | 0.955 |
| CUPP | 0.888 | 0.895 | 0.290 | 0.901 | 0.888 | 0.893 | 0.295 | 0.897 | 0.885 | 0.891 | 0.300 | 0.898 |

## 10.1 Testing for statistically significantly different performances between taxonomic kingdoms

## SI table 88: Output of two-way ANOVA testing for significant differences between the mean RI between CAZyme classifications and taxonomic kingdom of the test set

| Term | Degrees of Freedom | Sum of Squares | Mean Square | F-Statistic | P-value |
|---|---|---|---|---|---|
| Tax_group | 2 | 0.102294 | 0.051147 | 6.507632 | 0.001492 |
| Prediction_tool | 12 | 61.14061 | 5.09505 | 648.2651 | 0 |
| Tax_group:Prediction_tool | 24 | 0.433985 | 0.018083 | 2.300739 | 0.000293 |
| Residuals | 415961 | 3269.252 | 0.00786 | NA | NA |

**SI table 89: Output of two-way ANOVA testing for significant differences between the mean ARI between CAZyme classifications and taxonomic kingdom of the test set**

| Term | Degrees of Freedom | Sum of Squares | Mean Square | F-Statistic | P-value |
|---|---|---|---|---|---|
| Tax_group | 2 | 1.187905 | 0.593953 | 9.780972 | 5.65E-05 |
| Prediction_tool | 12 | 409.1572 | 34.09644 | 561.4863 | 0 |
| Tax_group:Prediction_tool | 24 | 3.818196 | 0.159092 | 2.619854 | 2.48E-05 |
| Residuals | 415961 | 25259.37 | 0.060725 | NA | NA |

## 10.2 Testing for statistically significant differences between tools

### SI table 90: Output of a Tukey HSD test following a one-way ANOVA, testing for statistically significant differences between the mean Rand Index of the CAZyme classifiers (overleaf)

A one-way ANOVA was performed for the Rand Index (RI), testing for statistically significant differences between the mean RI attained by each tool when pooling **all** test sets. A Tukey HSD test was performed to identify between which tools the RI was statistically significantly different.

| Classifier 1 | Classifier 2 | Mean difference | Lower 95% Confidence Interval | Upper 95% Confidence Interval | Adjusted P-value |
|---|---|---|---|---|---|
| dbCAN_2 | CUPP | 0.016970833 | 0.013687092 | 0.020254575 | 0 |
| dbCAN_2:DIAMOND | CUPP | 0.019254167 | 0.015970425 | 0.022537908 | 0 |
| dbCAN_2:Hotpep | CUPP | -0.013195833 | -0.016479575 | -0.009912092 | 0 |
| dbCAN_3 | CUPP | 0.024054167 | 0.020770425 | 0.027337908 | 0 |
| dbCAN_3:DIAMOND | CUPP | 0.027608333 | 0.024324592 | 0.030892075 | 0 |
| dbCAN_3:HMMER | CUPP | 0.0114875 | 0.008203759 | 0.014771241 | 0 |
| dbCAN_4 | CUPP | 0.0257125 | 0.022428759 | 0.028996241 | 0 |
| dbCAN_4:dbCANsub | CUPP | 0.025170833 | 0.021887092 | 0.028454575 | 0 |
| dbCAN_4:DIAMOND | CUPP | 0.028154167 | 0.024870425 | 0.031437908 | 0 |
| dbCAN_4:HMMER | CUPP | 0.012795833 | 0.009512092 | 0.016079575 | 0 |
| dbCAN_2:Hotpep | dbCAN_2 | -0.030166667 | -0.033450408 | -0.026882925 | 0 |
| dbCAN_3:eCAMI | dbCAN_2 | -0.015845833 | -0.019129575 | -0.012562092 | 0 |
| dbCAN_2:Hotpep | dbCAN_2:DIAMOND | -0.03245 | -0.035733741 | -0.029166259 | 0 |
| dbCAN_3:eCAMI | dbCAN_2:DIAMOND | -0.018129167 | -0.021412908 | -0.014845425 | 0 |
| dbCAN_2:Hotpep | dbCAN_2:HMMER | -0.022733333 | -0.026017075 | -0.019449592 | 0 |
| dbCAN_3 | dbCAN_2:HMMER | 0.014516667 | 0.011232925 | 0.017800408 | 0 |
| dbCAN_3:DIAMOND | dbCAN_2:HMMER | 0.018070833 | 0.014787092 | 0.021354575 | 0 |
| dbCAN_4 | dbCAN_2:HMMER | 0.016175 | 0.012891259 | 0.019458741 | 0 |
| dbCAN_4:dbCANsub | dbCAN_2:HMMER | 0.015633333 | 0.012349592 | 0.018917075 | 0 |
| dbCAN_4:DIAMOND | dbCAN_2:HMMER | 0.018616667 | 0.015332925 | 0.021900408 | 0 |
| dbCAN_3 | dbCAN_2:Hotpep | 0.03725 | 0.033966259 | 0.040533741 | 0 |
| dbCAN_3:DIAMOND | dbCAN_2:Hotpep | 0.040804167 | 0.037520425 | 0.044087908 | 0 |
| dbCAN_3:eCAMI | dbCAN_2:Hotpep | 0.014320833 | 0.011037092 | 0.017604575 | 0 |
| dbCAN_3:HMMER | dbCAN_2:Hotpep | 0.024683333 | 0.021399592 | 0.027967075 | 0 |
| dbCAN_4 | dbCAN_2:Hotpep | 0.038908333 | 0.035624592 | 0.042192075 | 0 |
| dbCAN_4:dbCANsub | dbCAN_2:Hotpep | 0.038366667 | 0.035082925 | 0.041650408 | 0 |
| dbCAN_4:DIAMOND | dbCAN_2:Hotpep | 0.04135 | 0.038066259 | 0.044633741 | 0 |
| dbCAN_4:HMMER | dbCAN_2:Hotpep | 0.025991667 | 0.022707925 | 0.029275408 | 0 |
| dbCAN_3:eCAMI | dbCAN_3 | -0.022929167 | -0.026212908 | -0.019645425 | 0 |
| dbCAN_3:HMMER | dbCAN_3 | -0.012566667 | -0.015850408 | -0.009282925 | 0 |
| dbCAN_4:HMMER | dbCAN_3 | -0.011258333 | -0.014542075 | -0.007974592 | 0 |
| dbCAN_3:eCAMI | dbCAN_3:DIAMOND | -0.026483333 | -0.029767075 | -0.023199592 | 0 |
| dbCAN_3:HMMER | dbCAN_3:DIAMOND | -0.016120833 | -0.019404575 | -0.012837092 | 0 |
| dbCAN_4:HMMER | dbCAN_3:DIAMOND | -0.0148125 | -0.018096241 | -0.011528759 | 0 |
| dbCAN_4 | dbCAN_3:eCAMI | 0.0245875 | 0.021303759 | 0.027871241 | 0 |
| dbCAN_4:dbCANsub | dbCAN_3:eCAMI | 0.024045833 | 0.020762092 | 0.027329575 | 0 |
| dbCAN_4:DIAMOND | dbCAN_3:eCAMI | 0.027029167 | 0.023745425 | 0.030312908 | 0 |
| dbCAN_4:HMMER | dbCAN_3:eCAMI | 0.011670833 | 0.008387092 | 0.014954575 | 0 |
| dbCAN_4 | dbCAN_3:HMMER | 0.014225 | 0.010941259 | 0.017508741 | 0 |
| dbCAN_4:dbCANsub | dbCAN_3:HMMER | 0.013683333 | 0.010399592 | 0.016967075 | 0 |
| dbCAN_4:DIAMOND | dbCAN_3:HMMER | 0.016666667 | 0.013382925 | 0.019950408 | 0 |
| dbCAN_4:HMMER | dbCAN_4 | -0.012916667 | -0.016200408 | -0.009632925 | 0 |
| dbCAN_4:HMMER | dbCAN_4:dbCANsub | -0.012375 | -0.015658741 | -0.009091259 | 0 |
| dbCAN_4:HMMER | dbCAN_4:DIAMOND | -0.015358333 | -0.018642075 | -0.012074592 | 0 |
| dbCAN_4:DIAMOND | dbCAN_2 | 0.011183333 | 0.007899592 | 0.014467075 | 2.02E-14 |
| dbCAN_4:DIAMOND | dbCAN_2:DIAMOND | 0.0089 | 0.005616259 | 0.012183741 | 8.85E-14 |
| dbCAN_2:HMMER | CUPP | 0.0095375 | 0.006253759 | 0.012821241 | 9.63E-14 |
| dbCAN_4 | dbCAN_2 | 0.008741667 | 0.005457925 | 0.012025408 | 9.80E-14 |
| dbCAN_3:eCAMI | dbCAN_2:HMMER | -0.0084125 | -0.011696241 | -0.005128759 | 1.25E-13 |
| dbCAN_3:HMMER | dbCAN_3:eCAMI | 0.0103625 | 0.007078759 | 0.013646241 | 1.25E-13 |
| dbCAN_3:DIAMOND | dbCAN_2 | 0.0106375 | 0.007353759 | 0.013921241 | 1.28E-13 |
| dbCAN_3:DIAMOND | dbCAN_2:DIAMOND | 0.008354167 | 0.005070425 | 0.011637908 | 1.33E-13 |
| dbCAN_2:HMMER | dbCAN_2:DIAMOND | -0.009716667 | -0.013000408 | -0.006432925 | 1.35E-13 |
| dbCAN_4:dbCANsub | dbCAN_2 | 0.0082 | 0.004916259 | 0.011483741 | 1.55E-13 |
| dbCAN_3:HMMER | dbCAN_2:DIAMOND | -0.007766667 | -0.011050408 | -0.004482925 | 5.58E-13 |
| dbCAN_2:HMMER | dbCAN_2 | -0.007433333 | -0.010717075 | -0.004149592 | 5.10E-12 |
| dbCAN_3 | dbCAN_2 | 0.007083333 | 0.003799592 | 0.010367075 | 6.98E-11 |
| dbCAN_4 | dbCAN_2:DIAMOND | 0.006458333 | 0.003174592 | 0.009742075 | 5.65E-09 |
| dbCAN_4:HMMER | dbCAN_2:DIAMOND | -0.006458333 | -0.009742075 | -0.003174592 | 5.65E-09 |
| dbCAN_4:dbCANsub | dbCAN_2:DIAMOND | 0.005916667 | 0.002632925 | 0.009200408 | 1.85E-07 |
| dbCAN_3:HMMER | dbCAN_2 | -0.005483333 | -0.008767075 | -0.002199592 | 2.45E-06 |
| dbCAN_3 | dbCAN_2:DIAMOND | 0.0048 | 0.001516259 | 0.008083741 | 9.63E-05 |
| dbCAN_4:HMMER | dbCAN_2 | -0.004175 | -0.007458741 | -0.000891259 | 0.001778215 |
| dbCAN_4:DIAMOND | dbCAN_3 | 0.0041 | 0.000816259 | 0.007383741 | 0.002447992 |
| dbCAN_3:DIAMOND | dbCAN_3 | 0.003554167 | 0.000270425 | 0.006837908 | 0.020307247 |
| dbCAN_4:HMMER | dbCAN_2:HMMER | 0.003258333 | -2.54E-05 | 0.006542075 | 0.05412477 |
| dbCAN_4:DIAMOND | dbCAN_4:dbCANsub | 0.002983333 | -0.000300408 | 0.006267075 | 0.119901623 |
| dbCAN_4:DIAMOND | dbCAN_4 | 0.002441667 | -0.000842075 | 0.005725408 | 0.39928434 |
| dbCAN_4:dbCANsub | dbCAN_3:DIAMOND | -0.0024375 | -0.005721241 | 0.000846241 | 0.402189424 |
| dbCAN_2:DIAMOND | dbCAN_2 | 0.002283333 | -0.001000408 | 0.005567075 | 0.514281808 |
| dbCAN_3:HMMER | dbCAN_2:HMMER | 0.00195 | -0.001333741 | 0.005233741 | 0.754551099 |
| dbCAN_4 | dbCAN_3:DIAMOND | -0.001895833 | -0.005179575 | 0.001387908 | 0.788395383 |
| dbCAN_4 | dbCAN_3 | 0.001658333 | -0.001625408 | 0.004942075 | 0.905654999 |
| dbCAN_4:HMMER | dbCAN_3:HMMER | 0.001308333 | -0.001975408 | 0.004592075 | 0.98432884 |
| dbCAN_3:eCAMI | CUPP | 0.001125 | -0.002158741 | 0.004408741 | 0.995884766 |
| dbCAN_4:dbCANsub | dbCAN_3 | 0.001116667 | -0.002167075 | 0.004400408 | 0.996159274 |
| dbCAN_4:DIAMOND | dbCAN_3:DIAMOND | 0.000545833 | -0.002737908 | 0.003829575 | 0.999997939 |
| dbCAN_4:dbCANsub | dbCAN_4 | -0.000541667 | -0.003825408 | 0.002742075 | 0.99999811 |

## SI table 91: Output of a Tukey HSD test following a one-way ANOVA, testing for statistically significant differences between the mean Adjusted Rand Index of the CAZyme classifiers (overleaf)

A one-way ANOVA was performed for the Adjusted Rand Index (ARI), testing for statistically significant differences between the mean RI attained by each tool when pooling **all** test sets. A Tukey HSD test was performed to identify between which tools the RI was statistically significantly different.

| Classifier 1 | Classifier 2 | Mean difference | Lower 95% Confidence Interval | Upper 95% Confidence Interval | Adjusted P-value |
|---|---|---|---|---|---|
| dbCAN_2 | CUPP | 0.040450773 | 0.031323032 | 0.049578515 | 0 |
| dbCAN_2:DIAMOND | CUPP | 0.047092416 | 0.037964674 | 0.056220157 | 0 |
| dbCAN_2:Hotpep | CUPP | -0.034078926 | -0.043206668 | -0.024951185 | 0 |
| dbCAN_3 | CUPP | 0.06092147 | 0.051793728 | 0.070049211 | 0 |
| dbCAN_3:DIAMOND | CUPP | 0.071157045 | 0.062029304 | 0.080284787 | 0 |
| dbCAN_4 | CUPP | 0.066299187 | 0.057171445 | 0.075426929 | 0 |
| dbCAN_4:dbCANsub | CUPP | 0.064953197 | 0.055825455 | 0.074080939 | 0 |
| dbCAN_4:DIAMOND | CUPP | 0.072445632 | 0.063317891 | 0.081573374 | 0 |
| dbCAN_2:Hotpep | dbCAN_2 | -0.0745297 | -0.083657442 | -0.065401958 | 0 |
| dbCAN_3:eCAMI | dbCAN_2 | -0.039317691 | -0.048445432 | -0.030189949 | 0 |
| dbCAN_4:DIAMOND | dbCAN_2 | 0.031994859 | 0.022867117 | 0.041122601 | 0 |
| dbCAN_2:Hotpep | dbCAN_2:DIAMOND | -0.081171342 | -0.090299084 | -0.0720436 | 0 |
| dbCAN_3:eCAMI | dbCAN_2:DIAMOND | -0.045959333 | -0.055087075 | -0.036831591 | 0 |
| dbCAN_2:Hotpep | dbCAN_2:HMMER | -0.056373633 | -0.065501375 | -0.047245892 | 0 |
| dbCAN_3 | dbCAN_2:HMMER | 0.038626763 | 0.029499021 | 0.047754504 | 0 |
| dbCAN_3:DIAMOND | dbCAN_2:HMMER | 0.048862338 | 0.039734597 | 0.05799008 | 0 |
| dbCAN_4 | dbCAN_2:HMMER | 0.04400448 | 0.034876738 | 0.053132222 | 0 |
| dbCAN_4:dbCANsub | dbCAN_2:HMMER | 0.04265849 | 0.033530748 | 0.051786232 | 0 |
| dbCAN_4:DIAMOND | dbCAN_2:HMMER | 0.050150925 | 0.041023184 | 0.059278667 | 0 |
| dbCAN_3 | dbCAN_2:Hotpep | 0.095000396 | 0.085872654 | 0.104128138 | 0 |
| dbCAN_3:DIAMOND | dbCAN_2:Hotpep | 0.105235972 | 0.09610823 | 0.114363714 | 0 |
| dbCAN_3:eCAMI | dbCAN_2:Hotpep | 0.035212009 | 0.026084267 | 0.044339751 | 0 |
| dbCAN_3:HMMER | dbCAN_2:Hotpep | 0.061984481 | 0.052856739 | 0.071112223 | 0 |
| dbCAN_4 | dbCAN_2:Hotpep | 0.100378113 | 0.091250372 | 0.109505855 | 0 |
| dbCAN_4:dbCANsub | dbCAN_2:Hotpep | 0.099032123 | 0.089904382 | 0.108159865 | 0 |
| dbCAN_4:DIAMOND | dbCAN_2:Hotpep | 0.106524559 | 0.097396817 | 0.1156523 | 0 |
| dbCAN_4:HMMER | dbCAN_2:Hotpep | 0.064941003 | 0.055813261 | 0.074068744 | 0 |
| dbCAN_3:eCAMI | dbCAN_3 | -0.059788387 | -0.068916129 | -0.050660645 | 0 |
| dbCAN_3:HMMER | dbCAN_3 | -0.033015915 | -0.042143657 | -0.023888173 | 0 |
| dbCAN_3:eCAMI | dbCAN_3:DIAMOND | -0.070023963 | -0.079151704 | -0.060896221 | 0 |
| dbCAN_3:HMMER | dbCAN_3:DIAMOND | -0.043251491 | -0.052379233 | -0.034123749 | 0 |
| dbCAN_4:HMMER | dbCAN_3:DIAMOND | -0.040294969 | -0.049422711 | -0.031167227 | 0 |
| dbCAN_4 | dbCAN_3:eCAMI | 0.065166104 | 0.056038363 | 0.074293846 | 0 |
| dbCAN_4:dbCANsub | dbCAN_3:eCAMI | 0.063820114 | 0.054692373 | 0.072947856 | 0 |
| dbCAN_4:DIAMOND | dbCAN_3:eCAMI | 0.07131255 | 0.062184808 | 0.080440291 | 0 |
| dbCAN_4 | dbCAN_3:HMMER | 0.038393632 | 0.029265891 | 0.047521374 | 0 |
| dbCAN_4:dbCANsub | dbCAN_3:HMMER | 0.037047642 | 0.027919901 | 0.046175384 | 0 |
| dbCAN_4:DIAMOND | dbCAN_3:HMMER | 0.044540078 | 0.035412336 | 0.05366782 | 0 |
| dbCAN_4:HMMER | dbCAN_4 | -0.035437111 | -0.044564852 | -0.026309369 | 0 |
| dbCAN_4:HMMER | dbCAN_4:dbCANsub | -0.034091121 | -0.043218862 | -0.024963379 | 0 |
| dbCAN_4:HMMER | dbCAN_4:DIAMOND | -0.041583556 | -0.050711298 | -0.032455814 | 0 |
| dbCAN_4:HMMER | CUPP | 0.030862076 | 0.021734335 | 0.039989818 | 3.18E-14 |
| dbCAN_3:DIAMOND | dbCAN_2 | 0.030706272 | 0.02157853 | 0.039834014 | 4.33E-14 |
| dbCAN_2:HMMER | dbCAN_2:DIAMOND | -0.024797709 | -0.03392545 | -0.015669967 | 8.60E-14 |
| dbCAN_4:dbCANsub | dbCAN_2 | 0.024502424 | 0.015374682 | 0.033630165 | 9.41E-14 |
| dbCAN_3:DIAMOND | dbCAN_2:DIAMOND | 0.02406463 | 0.014936888 | 0.033192372 | 1.04E-13 |
| dbCAN_4:HMMER | dbCAN_3 | -0.030059393 | -0.039187135 | -0.020931652 | 1.19E-13 |
| dbCAN_4:HMMER | dbCAN_3:eCAMI | 0.029728994 | 0.020601252 | 0.038856735 | 1.23E-13 |
| dbCAN_4 | dbCAN_2 | 0.025848414 | 0.016720672 | 0.034976155 | 1.24E-13 |
| dbCAN_3:HMMER | CUPP | 0.027905555 | 0.018777813 | 0.037033296 | 1.31E-13 |
| dbCAN_4:DIAMOND | dbCAN_2:DIAMOND | 0.025353217 | 0.016225475 | 0.034480959 | 1.49E-13 |
| dbCAN_3:HMMER | dbCAN_3:eCAMI | 0.026772472 | 0.01764473 | 0.035900214 | 1.51E-13 |
| dbCAN_2:HMMER | CUPP | 0.022294707 | 0.013166965 | 0.031422449 | 2.10E-13 |
| dbCAN_3:eCAMI | dbCAN_2:HMMER | -0.021161624 | -0.030289366 | -0.012033883 | 1.32E-12 |
| dbCAN_3 | dbCAN_2 | 0.020470696 | 0.011342954 | 0.029598438 | 8.59E-12 |
| dbCAN_4 | dbCAN_2:DIAMOND | 0.019206771 | 0.01007903 | 0.028334513 | 2.46E-10 |
| dbCAN_3:HMMER | dbCAN_2:DIAMOND | -0.019186861 | -0.028314603 | -0.010059119 | 2.59E-10 |
| dbCAN_2:HMMER | dbCAN_2 | -0.018156066 | -0.027283808 | -0.009028325 | 3.44E-09 |
| dbCAN_4:dbCANsub | dbCAN_2:DIAMOND | 0.017860781 | 0.00873304 | 0.026988523 | 7.04E-09 |
| dbCAN_4:HMMER | dbCAN_2:DIAMOND | -0.016230339 | -0.025358081 | -0.007102598 | 2.99E-07 |
| dbCAN_3 | dbCAN_2:DIAMOND | 0.013829054 | 0.004701312 | 0.022956796 | 3.94E-05 |
| dbCAN_3:HMMER | dbCAN_2 | -0.012545219 | -0.021672961 | -0.003417477 | 0.000387734 |
| dbCAN_4:DIAMOND | dbCAN_3 | 0.011524163 | 0.002396421 | 0.020651904 | 0.002014945 |
| dbCAN_3:DIAMOND | dbCAN_3 | 0.010235576 | 0.001107834 | 0.019363318 | 0.012792516 |
| dbCAN_4:HMMER | dbCAN_2 | -0.009588697 | -0.018716439 | -0.000460955 | 0.029127937 |
| dbCAN_4:HMMER | dbCAN_2:HMMER | 0.008567369 | -0.000560372 | 0.017695111 | 0.091316568 |
| dbCAN_4:DIAMOND | dbCAN_4:dbCANsub | 0.007492435 | -0.001635306 | 0.016620177 | 0.242075324 |
| dbCAN_2:DIAMOND | dbCAN_2 | 0.006641642 | -0.0024861 | 0.015769384 | 0.436322967 |
| dbCAN_4:dbCANsub | dbCAN_3:DIAMOND | -0.006203848 | -0.01533159 | 0.002923893 | 0.552868527 |
| dbCAN_4:DIAMOND | dbCAN_4 | 0.006146445 | -0.002981296 | 0.015274187 | 0.568372319 |
| dbCAN_3:HMMER | dbCAN_2:HMMER | 0.005610848 | -0.003516894 | 0.014738589 | 0.708740483 |
| dbCAN_4 | dbCAN_3 | 0.005377717 | -0.003750024 | 0.014505459 | 0.764368378 |
| dbCAN_4 | dbCAN_3:DIAMOND | -0.004857858 | -0.0139856 | 0.004269883 | 0.867959033 |
| dbCAN_4:dbCANsub | dbCAN_3 | 0.004031727 | -0.005096014 | 0.013159469 | 0.963985827 |
| dbCAN_4:HMMER | dbCAN_3:HMMER | 0.002956522 | -0.00617122 | 0.012084263 | 0.997571754 |
| dbCAN_4:dbCANsub | dbCAN_4 | -0.00134599 | -0.010473732 | 0.007781752 | 0.999999473 |
| dbCAN_4:DIAMOND | dbCAN_3:DIAMOND | 0.001288587 | -0.007839155 | 0.010416329 | 0.99999968 |
| dbCAN_3:eCAMI | CUPP | 0.001133083 | -0.007994659 | 0.010260824 | 0.999999928 |

## SI table 92: Output of Tukey HSD test following a two-way ANOVA, testing for statistically significant differences between the Rand Index (RI) and Adjusted Rand Index (ARI) attained by a tool between the bacterial and eukaryotic test set.

Reporting only results where a statistically significant difference is detected between the mean RI or ARI achieved for each taxonomic kingdom by the same classifier. The complete output is available in the online repository.

| Statistic | Kingdom of Classifier 1 | Classifier 1 | Kingdom of Classifier 2 | Classifier 2 | Mean Difference | Lower 95% Confidence Interval | Upper 95% Confidence Interval | Adjusted P-value |
|---|---|---|---|---|---|---|---|---|
| RI | Eukaryote | dbCAN_3:eCAMI | Bacteria | dbCAN_3:eCAMI | -0.00635 | -0.01178256 | -0.00091744 | 0.003656548 |
| RI | Eukaryote | dbCAN_4:sub | Bacteria | dbCAN_4:sub | -0.006575 | -0.01200756 | -0.00114244 | 0.001756852 |
| ARI | Eukaryote | dbCAN_3:eCAMI | Bacteria | dbCAN_3:eCAMI | -0.02416926 | -0.039269767 | -0.009068753 | 4.07E-07 |
| AIR | Eukaryote | dbCAN_4:sub | Bacteria | dbCAN_4:sub | -0.01661604 | -0.031716547 | -0.001515533 | 0.011398299 |

# 11 Overall performance of CAZy family classification (across all CAZy families and classes)

### SI Table 93: Summary of CAZy family classification (overleaf)

SI table 93 lists the statistical parameter values for the binary classification of CAZy families across all test sets. This treats the classification of each protein against each family as a separate decision, i.e. test for correct classification against GH1, then GH2, etc. The mean (standard deviation) and lower and upper 95% confidence intervals (CIs) were calculated for the statistical parameter values by pooling all test sets, and across all CAZy families, across all and per CAZy class.

| CAZy class | Classifier | | CUPP | dbCAN_2 | dbCAN_2:HMMER | dbCAN_2:DIAMOND | dbCAN_2:Hotpep | dbCAN_3 | dbCAN_3:HMMER | dbCAN_3:DIAMOND | dbCAN_3:eCAMI | dbCAN_4 | dbCAN_4:HMMER | dbCAN_4:DIAMOND | dbCAN_4:dbCAN-sub |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| All CAZy classes | Specificity | Mean | 1.000 | 1.000 | 1.000 | 1.000 | 0.999 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 |
| | | Standard Deviation | 0.000 | 0.000 | 0.000 | 0.000 | 0.002 | 0.000 | 0.000 | 0.000 | 0.001 | 0.000 | 0.000 | 0.000 | 0.000 |
| | | Lower CI | 1.000 | 1.000 | 1.000 | 1.000 | 0.999 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 |
| | | Upper CI | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 |
| | Sensitivity | Mean | 0.642 | 0.839 | 0.814 | 0.845 | 0.729 | 0.905 | 0.851 | 0.930 | 0.701 | 0.900 | 0.858 | 0.938 | 0.891 |
| | | Standard Deviation | 0.439 | 0.294 | 0.342 | 0.289 | 0.353 | 0.236 | 0.305 | 0.209 | 0.361 | 0.261 | 0.298 | 0.194 | 0.272 |
| | | Lower CI | 0.594 | 0.807 | 0.777 | 0.813 | 0.691 | 0.879 | 0.818 | 0.907 | 0.662 | 0.872 | 0.825 | 0.917 | 0.861 |
| | | Upper CI | 0.690 | 0.871 | 0.851 | 0.876 | 0.766 | 0.931 | 0.884 | 0.953 | 0.740 | 0.929 | 0.890 | 0.959 | 0.920 |
| | Precision | Mean | 0.679 | 0.876 | 0.829 | 0.874 | 0.723 | 0.914 | 0.863 | 0.913 | 0.738 | 0.891 | 0.870 | 0.923 | 0.881 |
| | | Standard Deviation | 0.452 | 0.297 | 0.340 | 0.289 | 0.396 | 0.240 | 0.302 | 0.223 | 0.389 | 0.271 | 0.293 | 0.206 | 0.280 |
| | | Lower CI | 0.630 | 0.843 | 0.792 | 0.842 | 0.681 | 0.888 | 0.830 | 0.889 | 0.696 | 0.861 | 0.838 | 0.901 | 0.851 |
| | | Upper CI | 0.729 | 0.908 | 0.866 | 0.905 | 0.766 | 0.940 | 0.896 | 0.938 | 0.780 | 0.920 | 0.901 | 0.946 | 0.912 |
| | F1-score Mean | Mean | 0.651 | 0.846 | 0.808 | 0.850 | 0.696 | 0.901 | 0.843 | 0.916 | 0.698 | 0.886 | 0.850 | 0.923 | 0.875 |
| | | Standard Deviation | 0.438 | 0.292 | 0.337 | 0.285 | 0.366 | 0.235 | 0.301 | 0.212 | 0.364 | 0.264 | 0.293 | 0.198 | 0.274 |
| | | Lower CI | 0.603 | 0.814 | 0.771 | 0.819 | 0.657 | 0.876 | 0.811 | 0.892 | 0.658 | 0.857 | 0.818 | 0.901 | 0.845 |
| | | Upper CI | 0.699 | 0.878 | 0.845 | 0.881 | 0.735 | 0.927 | 0.876 | 0.939 | 0.737 | 0.914 | 0.882 | 0.945 | 0.905 |
| | Accuracy | Mean | 0.999 | 0.999 | 0.999 | 1.000 | 0.999 | 1.000 | 0.999 | 1.000 | 0.999 | 1.000 | 0.999 | 1.000 | 1.000 |
| | | Standard Deviation | 0.002 | 0.001 | 0.002 | 0.001 | 0.003 | 0.001 | 0.002 | 0.001 | 0.002 | 0.001 | 0.002 | 0.001 | 0.001 |
| | | Lower CI | 0.999 | 0.999 | 0.999 | 0.999 | 0.998 | 1.000 | 0.999 | 1.000 | 0.999 | 1.000 | 0.999 | 1.000 | 1.000 |
| | | Upper CI | 0.999 | 1.000 | 1.000 | 1.000 | 0.999 | 1.000 | 1.000 | 1.000 | 0.999 | 1.000 | 1.000 | 1.000 | 1.000 |
| GH | Specificity | Mean | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 |
| | | Standard Deviation | 0.000 | 0.000 | 0.000 | 0.000 | 0.001 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |
| | | Lower CI | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 |
| | | Upper CI | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 |
| | Sensitivity | Mean | 0.785 | 0.867 | 0.848 | 0.868 | 0.782 | 0.928 | 0.880 | 0.944 | 0.726 | 0.920 | 0.887 | 0.949 | 0.910 |
| | | Standard Deviation | 0.364 | 0.270 | 0.313 | 0.270 | 0.320 | 0.201 | 0.279 | 0.193 | 0.354 | 0.236 | 0.268 | 0.175 | 0.250 |
| | | Lower CI | 0.723 | 0.821 | 0.796 | 0.822 | 0.728 | 0.894 | 0.833 | 0.911 | 0.666 | 0.880 | 0.842 | 0.919 | 0.868 |
| | | Upper CI | 0.847 | 0.913 | 0.901 | 0.914 | 0.836 | 0.962 | 0.926 | 0.977 | 0.785 | 0.960 | 0.932 | 0.979 | 0.952 |
| | Precision | Mean | 0.833 | 0.919 | 0.875 | 0.912 | 0.855 | 0.953 | 0.903 | 0.944 | 0.819 | 0.924 | 0.910 | 0.951 | 0.913 |
| | | Standard Deviation | 0.362 | 0.261 | 0.311 | 0.262 | 0.324 | 0.191 | 0.274 | 0.194 | 0.366 | 0.240 | 0.263 | 0.176 | 0.254 |
| | | Lower CI | 0.772 | 0.875 | 0.823 | 0.868 | 0.800 | 0.921 | 0.857 | 0.911 | 0.757 | 0.884 | 0.866 | 0.921 | 0.870 |
| | | Upper CI | 0.895 | 0.963 | 0.927 | 0.957 | 0.910 | 0.986 | 0.949 | 0.977 | 0.880 | 0.965 | 0.954 | 0.981 | 0.955 |
| | F1-score Mean | Mean | 0.800 | 0.888 | 0.855 | 0.885 | 0.809 | 0.938 | 0.885 | 0.943 | 0.760 | 0.920 | 0.893 | 0.948 | 0.908 |
| | | Standard Deviation | 0.360 | 0.262 | 0.310 | 0.262 | 0.315 | 0.193 | 0.275 | 0.191 | 0.353 | 0.235 | 0.264 | 0.173 | 0.248 |
| | | Lower CI | 0.739 | 0.844 | 0.804 | 0.841 | 0.755 | 0.905 | 0.839 | 0.910 | 0.701 | 0.880 | 0.848 | 0.919 | 0.866 |
| | | Upper CI | 0.862 | 0.933 | 0.907 | 0.930 | 0.862 | 0.971 | 0.932 | 0.975 | 0.820 | 0.960 | 0.937 | 0.978 | 0.950 |
| | Accuracy | Mean | 1.000 | 1.000 | 1.000 | 1.000 | 0.999 | 1.000 | 1.000 | 1.000 | 0.999 | 1.000 | 1.000 | 1.000 | 1.000 |
| | | Standard Deviation | 0.001 | 0.001 | 0.001 | 0.001 | 0.002 | 0.001 | 0.001 | 0.001 | 0.001 | 0.001 | 0.001 | 0.001 | 0.001 |
| | | Lower CI | 0.999 | 0.999 | 0.999 | 1.000 | 0.999 | 1.000 | 0.999 | 1.000 | 0.999 | 1.000 | 0.999 | 1.000 | 1.000 |
| | | Upper CI | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 |
| GT | Specificity | Mean | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 |
| | | Standard Deviation | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |
| | | Lower CI | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 |
| | | Upper CI | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 |
| | Sensitivity | Mean | 0.804 | 0.859 | 0.843 | 0.872 | 0.780 | 0.901 | 0.876 | 0.935 | 0.775 | 0.910 | 0.864 | 0.936 | 0.910 |
| | | Standard Deviation | 0.298 | 0.265 | 0.303 | 0.259 | 0.319 | 0.226 | 0.255 | 0.187 | 0.319 | 0.239 | 0.272 | 0.187 | 0.239 |
| | | Lower CI | 0.735 | 0.798 | 0.774 | 0.812 | 0.706 | 0.849 | 0.818 | 0.893 | 0.702 | 0.856 | 0.803 | 0.893 | 0.856 |
| | | Upper CI | 0.873 | 0.920 | 0.912 | 0.931 | 0.853 | 0.952 | 0.934 | 0.978 | 0.848 | 0.965 | 0.926 | 0.979 | 0.965 |
| | Precision | Mean | 0.878 | 0.910 | 0.843 | 0.910 | 0.863 | 0.924 | 0.876 | 0.936 | 0.867 | 0.895 | 0.865 | 0.936 | 0.894 |
| | | Standard Deviation | 0.293 | 0.252 | 0.307 | 0.252 | 0.309 | 0.227 | 0.259 | 0.199 | 0.307 | 0.251 | 0.276 | 0.198 | 0.251 |
| | | Lower CI | 0.811 | 0.852 | 0.773 | 0.852 | 0.791 | 0.872 | 0.817 | 0.890 | 0.797 | 0.838 | 0.802 | 0.891 | 0.837 |
| | | Upper CI | 0.946 | 0.968 | 0.913 | 0.968 | 0.934 | 0.976 | 0.935 | 0.981 | 0.937 | 0.953 | 0.927 | 0.981 | 0.951 |
| | F1-score Mean | Mean | 0.823 | 0.866 | 0.829 | 0.874 | 0.800 | 0.901 | 0.862 | 0.927 | 0.798 | 0.892 | 0.850 | 0.926 | 0.892 |
| | | Standard Deviation | 0.288 | 0.261 | 0.298 | 0.259 | 0.307 | 0.223 | 0.250 | 0.190 | 0.310 | 0.239 | 0.267 | 0.188 | 0.240 |
| | | Lower CI | 0.757 | 0.806 | 0.761 | 0.815 | 0.730 | 0.850 | 0.804 | 0.883 | 0.728 | 0.838 | 0.790 | 0.883 | 0.837 |
| | | Upper CI | 0.889 | 0.926 | 0.897 | 0.934 | 0.871 | 0.952 | 0.919 | 0.970 | 0.869 | 0.947 | 0.911 | 0.969 | 0.946 |
| | Accuracy | Mean | 0.999 | 0.999 | 0.999 | 0.999 | 0.999 | 1.000 | 0.999 | 1.000 | 0.999 | 1.000 | 0.999 | 1.000 | 1.000 |
| | | Standard Deviation | 0.002 | 0.002 | 0.002 | 0.001 | 0.005 | 0.001 | 0.002 | 0.001 | 0.002 | 0.001 | 0.002 | 0.001 | 0.001 |
| | | Lower CI | 0.999 | 0.999 | 0.999 | 0.999 | 0.998 | 0.999 | 0.999 | 1.000 | 0.999 | 0.999 | 0.999 | 1.000 | 0.999 |
| | | Upper CI | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 |
| PL | Specificity | Mean | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 |
| | | Standard Deviation | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |
| | | Lower CI | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 |
| | | Upper CI | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 |
| | Sensitivity | Mean | 0.705 | 0.808 | 0.836 | 0.848 | 0.695 | 0.956 | 0.923 | 0.975 | 0.660 | 0.923 | 0.923 | 0.978 | 0.924 |
| | | Standard Deviation | 0.416 | 0.345 | 0.350 | 0.341 | 0.365 | 0.208 | 0.231 | 0.104 | 0.394 | 0.231 | 0.231 | 0.104 | 0.232 |
| | | Lower CI | 0.525 | 0.659 | 0.685 | 0.700 | 0.541 | 0.866 | 0.823 | 0.930 | 0.489 | 0.823 | 0.823 | 0.933 | 0.824 |
| | | Upper CI | 0.885 | 0.957 | 0.988 | 0.995 | 0.849 | 1.046 | 1.023 | 1.020 | 0.830 | 1.023 | 1.023 | 1.023 | 1.024 |
| | Precision | Mean | 0.780 | 0.869 | 0.847 | 0.854 | 0.833 | 0.921 | 0.921 | 0.933 | 0.779 | 0.921 | 0.921 | 0.933 | 0.921 |
| | | Standard Deviation | 0.420 | 0.344 | 0.351 | 0.345 | 0.380 | 0.229 | 0.229 | 0.164 | 0.420 | 0.229 | 0.229 | 0.164 | 0.229 |
| | | Lower CI | 0.598 | 0.720 | 0.695 | 0.705 | 0.672 | 0.822 | 0.822 | 0.862 | 0.597 | 0.822 | 0.822 | 0.862 | 0.822 |
| | | Upper CI | 0.961 | 1.018 | 0.999 | 1.004 | 0.993 | 1.020 | 1.020 | 1.004 | 0.960 | 1.020 | 1.020 | 1.004 | 1.020 |
| | F1-score Mean | Mean | 0.731 | 0.832 | 0.834 | 0.848 | 0.747 | 0.934 | 0.914 | 0.948 | 0.704 | 0.914 | 0.914 | 0.949 | 0.914 |
| | | Standard Deviation | 0.409 | 0.339 | 0.344 | 0.340 | 0.361 | 0.216 | 0.222 | 0.131 | 0.396 | 0.222 | 0.222 | 0.131 | 0.222 |
| | | Lower CI | 0.554 | 0.685 | 0.685 | 0.701 | 0.594 | 0.841 | 0.817 | 0.891 | 0.533 | 0.817 | 0.817 | 0.892 | 0.818 |
| | | Upper CI | 0.908 | 0.979 | 0.983 | 0.995 | 0.899 | 1.028 | 1.010 | 1.004 | 0.875 | 1.010 | 1.010 | 1.006 | 1.010 |
| | Accuracy | Mean | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 |
| | | Standard Deviation | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |
| | | Lower CI | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 |
| | | Upper CI | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 |
| CE | Specificity | Mean | 1.000 | 1.000 | 1.000 | 1.000 | 0.999 | 1.000 | 1.000 | 1.000 | 0.999 | 1.000 | 1.000 | 1.000 | 1.000 |
| | | Standard Deviation | 0.001 | 0.001 | 0.001 | 0.001 | 0.001 | 0.001 | 0.001 | 0.001 | 0.001 | 0.001 | 0.001 | 0.001 | 0.001 |
| | | Lower CI | 0.999 | 0.999 | 0.999 | 0.999 | 0.999 | 0.999 | 0.999 | 0.999 | 0.999 | 0.999 | 0.999 | 0.999 | 0.999 |
| | | Upper CI | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 |
| | Sensitivity | Mean | 0.741 | 0.733 | 0.716 | 0.689 | 0.711 | 0.851 | 0.811 | 0.881 | 0.654 | 0.877 | 0.867 | 0.971 | 0.879 |
| | | Standard Deviation | 0.426 | 0.382 | 0.441 | 0.360 | 0.347 | 0.292 | 0.375 | 0.238 | 0.321 | 0.320 | 0.317 | 0.047 | 0.320 |
| | | Lower CI | 0.523 | 0.543 | 0.504 | 0.510 | 0.533 | 0.700 | 0.625 | 0.758 | 0.489 | 0.718 | 0.709 | 0.946 | 0.719 |
| | | Upper CI | 0.960 | 0.923 | 0.929 | 0.868 | 0.890 | 1.001 | 0.998 | 1.003 | 0.819 | 1.036 | 1.025 | 0.995 | 1.038 |
| | Precision | Mean | 0.669 | 0.682 | 0.602 | 0.710 | 0.697 | 0.801 | 0.691 | 0.804 | 0.708 | 0.763 | 0.751 | 0.875 | 0.761 |
| | | Standard Deviation | 0.431 | 0.392 | 0.420 | 0.390 | 0.348 | 0.320 | 0.381 | 0.301 | 0.372 | 0.342 | 0.341 | 0.216 | 0.341 |
| | | Lower CI | 0.447 | 0.487 | 0.399 | 0.516 | 0.519 | 0.636 | 0.501 | 0.649 | 0.517 | 0.593 | 0.581 | 0.764 | 0.592 |
| | | Upper CI | 0.890 | 0.878 | 0.804 | 0.904 | 0.876 | 0.965 | 0.880 | 0.959 | 0.900 | 0.932 | 0.921 | 0.986 | 0.931 |
| | F1-score Mean | Mean | 0.688 | 0.695 | 0.640 | 0.691 | 0.680 | 0.813 | 0.731 | 0.825 | 0.667 | 0.801 | 0.791 | 0.904 | 0.802 |
| | | Standard Deviation | 0.417 | 0.377 | 0.417 | 0.365 | 0.322 | 0.299 | 0.367 | 0.264 | 0.332 | 0.323 | 0.322 | 0.150 | 0.323 |
| | | Lower CI | 0.473 | 0.508 | 0.439 | 0.509 | 0.515 | 0.660 | 0.548 | 0.689 | 0.496 | 0.641 | 0.630 | 0.827 | 0.641 |
| | | Upper CI | 0.903 | 0.882 | 0.841 | 0.872 | 0.846 | 0.967 | 0.914 | 0.960 | 0.838 | 0.962 | 0.951 | 0.981 | 0.962 |
| | Accuracy | Mean | 0.999 | 0.999 | 0.999 | 0.999 | 0.999 | 0.999 | 0.999 | 0.999 | 0.999 | 0.999 | 0.999 | 0.999 | 0.999 |
| | | Standard Deviation | 0.001 | 0.001 | 0.001 | 0.001 | 0.002 | 0.002 | 0.001 | 0.001 | 0.002 | 0.001 | 0.001 | 0.001 | 0.001 |
| | | Lower CI | 0.999 | 0.998 | 0.999 | 0.998 | 0.998 | 0.998 | 0.999 | 0.999 | 0.998 | 0.999 | 0.999 | 0.999 | 0.999 |
| | | Upper CI | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 |
| AA | Specificity | Mean | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 |
| | | Standard Deviation | 0.001 | 0.001 | 0.001 | 0.001 | 0.001 | 0.001 | 0.001 | 0.001 | 0.001 | 0.001 | 0.001 | 0.001 | 0.001 |
| | | Lower CI | 0.999 | 0.999 | 0.999 | 0.999 | 0.999 | 0.999 | 0.999 | 0.999 | 0.999 | 0.999 | 0.999 | 0.999 | 0.999 |
| | | Upper CI | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 |
| | Sensitivity | Mean | 0.694 | 0.771 | 0.864 | 0.775 | 0.710 | 0.914 | 0.908 | 0.952 | 0.644 | 0.910 | 0.908 | 0.957 | 0.909 |
| | | Standard Deviation | 0.438 | 0.332 | 0.266 | 0.280 | 0.370 | 0.250 | 0.257 | 0.108 | 0.397 | 0.257 | 0.257 | 0.104 | 0.257 |
| | | Lower CI | 0.460 | 0.594 | 0.722 | 0.626 | 0.513 | 0.781 | 0.772 | 0.894 | 0.433 | 0.773 | 0.772 | 0.901 | 0.773 |
| | | Upper CI | 0.928 | 0.948 | 1.006 | 0.924 | 0.908 | 1.047 | 1.045 | 1.009 | 0.856 | 1.047 | 1.045 | 1.012 | 1.046 |
| | Precision | Mean | 0.651 | 0.797 | 0.825 | 0.847 | 0.721 | 0.850 | 0.832 | 0.856 | 0.662 | 0.838 | 0.832 | 0.855 | 0.835 |
| | | Standard Deviation | 0.432 | 0.337 | 0.279 | 0.259 | 0.392 | 0.257 | 0.271 | 0.170 | 0.414 | 0.267 | 0.271 | 0.179 | 0.269 |
| | | Lower CI | 0.421 | 0.617 | 0.676 | 0.709 | 0.512 | 0.713 | 0.688 | 0.765 | 0.442 | 0.696 | 0.688 | 0.760 | 0.691 |
| | | Upper CI | 0.881 | 0.977 | 0.974 | 0.984 | 0.930 | 0.987 | 0.977 | 0.946 | 0.883 | 0.981 | 0.977 | 0.951 | 0.978 |
| | F1-score Mean | Mean | 0.668 | 0.772 | 0.834 | 0.794 | 0.707 | 0.875 | 0.858 | 0.895 | 0.646 | 0.863 | 0.858 | 0.897 | 0.860 |
| | | Standard Deviation | 0.432 | 0.320 | 0.259 | 0.248 | 0.371 | 0.244 | 0.249 | 0.128 | 0.395 | 0.248 | 0.249 | 0.137 | 0.249 |
| | | Lower CI | 0.438 | 0.601 | 0.696 | 0.662 | 0.509 | 0.745 | 0.725 | 0.826 | 0.436 | 0.730 | 0.725 | 0.824 | 0.727 |
| | | Upper CI | 0.899 | 0.943 | 0.973 | 0.926 | 0.904 | 1.005 | 0.991 | 0.963 | 0.857 | 0.995 | 0.991 | 0.970 | 0.993 |
| | Accuracy | Mean | 0.999 | 0.999 | 0.999 | 0.999 | 0.999 | 1.000 | 1.000 | 1.000 | 0.999 | 1.000 | 1.000 | 1.000 | 1.000 |
| | | Standard Deviation | 0.001 | 0.001 | 0.001 | 0.001 | 0.001 | 0.001 | 0.001 | 0.001 | 0.001 | 0.001 | 0.001 | 0.001 | 0.001 |
| | | Lower CI | 0.999 | 0.999 | 0.999 | 0.999 | 0.999 | 0.999 | 0.999 | 0.999 | 0.999 | 0.999 | 0.999 | 0.999 | 0.999 |
| | | Upper CI | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 |
| CBM | Specificity | Mean | 1.000 | 1.000 | 1.000 | 1.000 | 0.998 | 1.000 | 1.000 | 1.000 | 0.999 | 1.000 | 1.000 | 1.000 | 1.000 |
| | | Standard Deviation | 0.000 | 0.000 | 0.000 | 0.000 | 0.004 | 0.000 | 0.000 | 0.000 | 0.002 | 0.000 | 0.000 | 0.000 | 0.001 |
| | | Lower CI | 1.000 | 1.000 | 1.000 | 1.000 | 0.997 | 1.000 | 1.000 | 1.000 | 0.999 | 1.000 | 1.000 | 1.000 | 1.000 |
| | | Upper CI | 1.000 | 1.000 | 1.000 | 1.000 | 0.999 | 1.000 | 1.000 | 1.000 | 0.999 | 1.000 | 1.000 | 1.000 | 1.000 |
| | Sensitivity | Mean | 0.000 | 0.810 | 0.696 | 0.820 | 0.588 | 0.847 | 0.714 | 0.877 | 0.605 | 0.833 | 0.731 | 0.881 | 0.801 |
| | | Standard Deviation | 0.000 | 0.322 | 0.412 | 0.317 | 0.409 | 0.305 | 0.403 | 0.297 | 0.397 | 0.334 | 0.392 | 0.294 | 0.353 |
| | | Lower CI | 0.000 | 0.723 | 0.586 | 0.735 | 0.490 | 0.764 | 0.606 | 0.797 | 0.506 | 0.744 | 0.626 | 0.802 | 0.707 |
| | | Upper CI | 0.000 | 0.897 | 0.807 | 0.906 | 0.686 | 0.929 | 0.822 | 0.957 | 0.703 | 0.923 | 0.836 | 0.961 | 0.896 |
| | Precision | Mean | 0.000 | 0.813 | 0.768 | 0.798 | 0.279 | 0.855 | 0.786 | 0.849 | 0.429 | 0.845 | 0.803 | 0.868 | 0.821 |
| | | Standard Deviation | 0.000 | 0.339 | 0.405 | 0.322 | 0.301 | 0.314 | 0.393 | 0.298 | 0.353 | 0.341 | 0.379 | 0.281 | 0.361 |
| | | Lower CI | 0.000 | 0.721 | 0.659 | 0.711 | 0.207 | 0.770 | 0.680 | 0.769 | 0.341 | 0.754 | 0.701 | 0.792 | 0.725 |
| | | Upper CI | 0.000 | 0.904 | 0.876 | 0.885 | 0.351 | 0.940 | 0.891 | 0.930 | 0.516 | 0.936 | 0.904 | 0.944 | 0.918 |
| | F1-score Mean | Mean | 0.000 | 0.791 | 0.700 | 0.797 | 0.342 | 0.833 | 0.718 | 0.855 | 0.466 | 0.814 | 0.735 | 0.858 | 0.783 |
| | | Standard Deviation | 0.000 | 0.326 | 0.407 | 0.312 | 0.316 | 0.306 | 0.397 | 0.293 | 0.345 | 0.341 | 0.386 | 0.292 | 0.359 |
| | | Lower CI | 0.000 | 0.703 | 0.591 | 0.713 | 0.266 | 0.751 | 0.611 | 0.776 | 0.380 | 0.723 | 0.631 | 0.779 | 0.687 |
| | | Upper CI | 0.000 | 0.879 | 0.809 | 0.882 | 0.417 | 0.916 | 0.824 | 0.934 | 0.551 | 0.906 | 0.838 | 0.937 | 0.879 |
| | Accuracy | Mean | 0.998 | 0.999 | 0.999 | 0.999 | 0.997 | 0.999 | 0.999 | 1.000 | 0.998 | 0.999 | 0.999 | 1.000 | 0.999 |
| | | Standard Deviation | 0.004 | 0.002 | 0.003 | 0.001 | 0.004 | 0.001 | 0.003 | 0.001 | 0.002 | 0.001 | 0.003 | 0.001 | 0.001 |
| | | Lower CI | 0.997 | 0.999 | 0.998 | 0.999 | 0.996 | 0.999 | 0.998 | 0.999 | 0.998 | 0.999 | 0.998 | 0.999 | 0.999 |
| | | Upper CI | 0.999 | 1.000 | 1.000 | 1.000 | 0.998 | 1.000 | 1.000 | 1.000 | 0.999 | 1.000 | 1.000 | 1.000 | 1.000 |

### 11.1 Testing for statistically significant differences between the tool

**SI Table 94: Output of Tukey HSD test, testing for statistically significant differences between the mean F1-score of tools**

| Classifier 1 | Classifier 2 | Mean Difference | Lower 95% Confidence Interval | Upper 95% Confidence Interval | Adjusted P-value |
|---|---|---|---|---|---|
| dbCAN_4:HMMER | dbCAN_2:Hotpep | -0.271878334 | -0.351400056 | -0.192356612 | 3.75E-08 |
| dbCAN_3:eCAMI | dbCAN_3:DIAMOND | -0.264615549 | -0.344137271 | -0.185093827 | 3.75E-08 |
| dbCAN_2:Hotpep | dbCAN_2:HMMER | 0.205714084 | 0.127200207 | 0.284227961 | 3.75E-08 |
| dbCAN_4 | dbCAN_3:eCAMI | 0.225329623 | 0.146586707 | 0.304072538 | 3.75E-08 |
| dbCAN_3:eCAMI | dbCAN_3 | -0.203835556 | -0.282578472 | -0.12509264 | 3.75E-08 |
| dbCAN_3:eCAMI | dbCAN_2 | 0.22720815 | 0.148694273 | 0.305722027 | 3.75E-08 |
| dbCAN_3:eCAMI | dbCAN_2:HMMER | -0.250384267 | -0.329905989 | -0.170862545 | 3.75E-08 |
| dbCAN_4:DIAMOND | dbCAN_3:HMMER | -0.223895575 | -0.303113206 | -0.144677944 | 3.75E-08 |
| dbCAN_4:DIAMOND | dbCAN_3:eCAMI | -0.234543954 | -0.31382175 | -0.155266158 | 3.75E-08 |
| dbCAN_3 | dbCAN_2:HMMER | 0.219945365 | 0.141431488 | 0.298459242 | 3.75E-08 |
| dbCAN_3:eCAMI | dbCAN_3:HMMER | -0.218066838 | -0.296809753 | -0.139323922 | 3.75E-08 |
| dbCAN_4:HMMER | dbCAN_3:eCAMI | -0.199014724 | -0.278112988 | -0.119916459 | 3.75E-08 |
| CUPP | dbCAN_3 | -0.19863028 | -0.278090525 | -0.119170036 | 3.75E-08 |
| CUPP | dbCAN_2 | -0.194781553 | -0.274241797 | -0.115321308 | 3.75E-08 |
| dbCAN_4:dbCANsub | dbCAN_3:eCAMI | -0.192406125 | -0.271563913 | -0.113248337 | 3.75E-08 |
| dbCAN_3:DIAMOND | dbCAN_2:HMMER | 0.18987377 | 0.111606961 | 0.26814058 | 3.75E-08 |
| dbCAN_3:HMMER | dbCAN_2:Hotpep | 0.187995243 | 0.109498673 | 0.266491812 | 3.75E-08 |
| dbCAN_3:DIAMOND | dbCAN_2:Hotpep | 0.179225392 | 0.101019524 | 0.257431259 | 3.75E-08 |
| dbCAN_4 | dbCAN_2:Hotpep | 0.177346864 | 0.098911058 | 0.255782669 | 3.75E-08 |
| CUPP | dbCAN_4 | -0.156900927 | -0.235939984 | -0.07786187 | 4.16E-08 |
| dbCAN_3:eCAMI | dbCAN_2:DIAMOND | 0.15434454 | 0.076259586 | 0.232429494 | 4.24E-08 |
| CUPP | dbCAN_3:DIAMOND | -0.153960097 | -0.232411706 | -0.075508487 | 4.42E-08 |
| dbCAN_4:dbCANsub | dbCAN_2:Hotpep | 0.152466012 | 0.074150765 | 0.230781259 | 4.70E-08 |
| CUPP | dbCAN_3:HMMER | -0.152081569 | -0.230762399 | -0.073400739 | 5.03E-08 |
| CUPP | dbCAN_2:DIAMOND | -0.150111369 | -0.228562979 | -0.071659759 | 5.69E-08 |
| dbCAN_3 | dbCAN_2:Hotpep | 0.147735941 | 0.069590691 | 0.225881191 | 6.91E-08 |
| CUPP | dbCAN_2:HMMER | -0.148232841 | -0.226913671 | -0.069552012 | 7.38E-08 |
| dbCAN_4:DIAMOND | dbCAN_2:HMMER | -0.145857414 | -0.224232779 | -0.067482048 | 9.61E-08 |
| dbCAN_2:Hotpep | dbCAN_2:DIAMOND | 0.114977407 | 0.036000588 | 0.193954225 | 0.000108294 |
| CUPP | dbCAN_4:dbCANsub | -0.112230744 | -0.190255721 | -0.034205766 | 0.000143682 |
| dbCAN_2:Hotpep | dbCAN_2 | -0.110352216 | -0.188607663 | -0.032096769 | 0.000225416 |
| CUPP | dbCAN_4:HMMER | 0.107714622 | 0.028737803 | 0.18669144 | 0.000462034 |
| CUPP | dbCAN_4:DIAMOND | 0.09348334 | 0.014506522 | 0.172460158 | 0.005877471 |
| dbCAN_4:DIAMOND | dbCAN_2:Hotpep | 0.079472209 | 0.000376566 | 0.158567852 | 0.04760133 |
| dbCAN_4 | dbCAN_2:HMMER | 0.077643027 | -0.001088177 | 0.15637423 | 0.057553851 |
| dbCAN_4:DIAMOND | dbCAN_2 | 0.077096781 | -0.002301555 | 0.156495117 | 0.066926913 |
| dbCAN_4:DIAMOND | dbCAN_2:DIAMOND | 0.073248054 | -0.006150282 | 0.15264639 | 0.105700701 |
| dbCAN_4:DIAMOND | dbCAN_4:HMMER | 0.07286361 | -0.006172462 | 0.151899683 | 0.106308351 |
| dbCAN_3:DIAMOND | dbCAN_3:HMMER | 0.072209424 | -0.006886219 | 0.151305067 | 0.115175414 |
| dbCAN_3:DIAMOND | dbCAN_2 | 0.069833996 | -0.00956434 | 0.149232332 | 0.15334392 |
| dbCAN_4:dbCANsub | dbCAN_2:HMMER | 0.066994648 | -0.011675973 | 0.145665269 | 0.191926743 |
| dbCAN_3:DIAMOND | dbCAN_2:DIAMOND | 0.065985269 | -0.013413067 | 0.145383605 | 0.224262939 |
| dbCAN_4:HMMER | dbCAN_3:DIAMOND | -0.065600825 | -0.144636898 | 0.013435247 | 0.226018201 |
| dbCAN_3:HMMER | dbCAN_3 | -0.057978142 | -0.137073786 | 0.021117501 | 0.422937568 |
| dbCAN_3 | dbCAN_2 | 0.055602715 | -0.023795621 | 0.135001051 | 0.501312227 |
| dbCAN_3 | dbCAN_2:DIAMOND | 0.051753987 | -0.027644349 | 0.131152323 | 0.620305862 |
| dbCAN_4:HMMER | dbCAN_3 | -0.051369544 | -0.130405616 | 0.027666529 | 0.624860069 |
| dbCAN_4:dbCANsub | dbCAN_4:DIAMOND | -0.047982759 | -0.127138292 | 0.031172774 | 0.726965694 |
| CUPP | dbCAN_3:eCAMI | -0.046548711 | -0.12535405 | 0.032256628 | 0.760392057 |
| CUPP | dbCAN_2:Hotpep | -0.044670184 | -0.123246666 | 0.033906299 | 0.804825134 |
| dbCAN_4:HMMER | dbCAN_2:HMMER | 0.042113796 | -0.036436627 | 0.12066422 | 0.861241415 |
| dbCAN_4 | dbCAN_3:HMMER | 0.042137829 | -0.03671257 | 0.120988228 | 0.863962434 |
| dbCAN_2:DIAMOND | dbCAN_2:HMMER | 0.041729353 | -0.037185563 | 0.120644269 | 0.872547185 |
| dbCAN_4:dbCANsub | dbCAN_3:DIAMOND | -0.040719974 | -0.119875507 | 0.038435559 | 0.892974492 |
| dbCAN_4 | dbCAN_2 | 0.039762401 | -0.039391628 | 0.118916431 | 0.908456252 |
| dbCAN_2:HMMER | dbCAN_2 | -0.037880625 | -0.116795542 | 0.041034291 | 0.932999159 |
| dbCAN_4:DIAMOND | dbCAN_4 | 0.03733438 | -0.041881364 | 0.116550125 | 0.941206493 |
| dbCAN_4 | dbCAN_2:DIAMOND | 0.035913674 | -0.043240356 | 0.115067703 | 0.955542366 |
| dbCAN_3:HMMER | dbCAN_2:HMMER | 0.035505198 | -0.043105164 | 0.11411556 | 0.957035975 |
| dbCAN_4:HMMER | dbCAN_4 | -0.03552923 | -0.114319873 | 0.043261412 | 0.957555626 |
| dbCAN_4:dbCANsub | dbCAN_3:HMMER | 0.03148945 | -0.047300458 | 0.110279358 | 0.983792408 |
| dbCAN_4 | dbCAN_3:DIAMOND | -0.030071595 | -0.109287339 | 0.04914415 | 0.98955844 |
| dbCAN_4:dbCANsub | dbCAN_2 | 0.029114022 | -0.049979748 | 0.108207793 | 0.99203614 |
| dbCAN_4:dbCANsub | dbCAN_3 | -0.026488692 | -0.105644225 | 0.052666841 | 0.996663854 |
| dbCAN_4:dbCANsub | dbCAN_2:DIAMOND | 0.025265295 | -0.053828476 | 0.104359066 | 0.997853578 |
| dbCAN_4:dbCANsub | dbCAN_4:HMMER | 0.024880852 | -0.053849254 | 0.103610958 | 0.998063848 |
| dbCAN_4:DIAMOND | dbCAN_3 | 0.021494067 | -0.057965795 | 0.100953928 | 0.999588061 |
| dbCAN_4 | dbCAN_3 | -0.015840313 | -0.095056058 | 0.063375431 | 0.99998355 |
| dbCAN_3:DIAMOND | dbCAN_3 | 0.014231282 | -0.06522858 | 0.093691143 | 0.999995164 |
| dbCAN_4:dbCANsub | dbCAN_4 | -0.010648379 | -0.089558853 | 0.068262096 | 0.999999808 |
| dbCAN_4:DIAMOND | dbCAN_3:DIAMOND | 0.007262785 | -0.072197076 | 0.086722646 | 0.999999998 |
| dbCAN_4:HMMER | dbCAN_3:HMMER | 0.006608599 | -0.072061294 | 0.085278491 | 0.999999999 |
| dbCAN_3:HMMER | dbCAN_2:DIAMOND | -0.006224155 | -0.08525799 | 0.072809679 | 1 |
| dbCAN_4:HMMER | dbCAN_2 | 0.004233171 | -0.074741046 | 0.083207388 | 1 |
| dbCAN_2:DIAMOND | dbCAN_2 | 0.003848728 | -0.075488036 | 0.083185491 | 1 |
| dbCAN_3:HMMER | dbCAN_2 | -0.002375428 | -0.081409262 | 0.076658407 | 1 |
| dbCAN_4:HMMER | dbCAN_2:DIAMOND | 0.000384443 | -0.078589774 | 0.07935866 | 1 |
| dbCAN_3:eCAMI | dbCAN_2:Hotpep | 0.001878528 | -0.075909685 | 0.079666741 | 1 |

**SI Table 95: Output of Tukey HSD test, testing for statistically significant differences between the mean accuracy of tools**

| Classifier 1 | Classifier 2 | Mean Difference | Lower 95% Confidence Interval | Upper 95% Confidence Interval | Adjusted P-value |
|---|---|---|---|---|---|
| dbCAN_4:DIAMOND | dbCAN_2:Hotpep | 0.001013532 | 0.000578986 | 0.001448077 | 3.75E-08 |
| dbCAN_3:DIAMOND | dbCAN_2:Hotpep | 0.000997444 | 0.000562898 | 0.00143199 | 3.75E-08 |
| dbCAN_4 | dbCAN_2:Hotpep | 0.000925227 | 0.000492049 | 0.001358406 | 3.76E-08 |
| dbCAN_4:dbCANsub | dbCAN_2:Hotpep | 0.000916507 | 0.000483666 | 0.001349348 | 3.77E-08 |
| dbCAN_3 | dbCAN_2:Hotpep | 0.000916452 | 0.000481907 | 0.001350998 | 3.77E-08 |
| dbCAN_2:Hotpep | dbCAN_2:DIAMOND | -0.000833921 | -0.001268122 | -0.00039972 | 5.42E-08 |
| dbCAN_2:Hotpep | dbCAN_2 | -0.000743429 | -0.001177631 | -0.000309228 | 1.18E-06 |
| dbCAN_4:HMMER | dbCAN_2:Hotpep | 0.000681135 | 0.000248963 | 0.001113306 | 1.41E-05 |
| dbCAN_3:HMMER | dbCAN_2:Hotpep | 0.00065691 | 0.000224405 | 0.001089416 | 3.79E-05 |
| dbCAN_2:Hotpep | dbCAN_2:HMMER | -0.000623994 | -0.001055834 | -0.000192154 | 0.000129216 |
| dbCAN_4:DIAMOND | dbCAN_3:eCAMI | 0.000585394 | 0.000149581 | 0.001021207 | 0.00063125 |
| dbCAN_3:eCAMI | dbCAN_3:DIAMOND | -0.000569306 | -0.00100512 | -0.000133493 | 0.001088439 |
| CUPP | dbCAN_4:DIAMOND | -0.000551196 | -0.00099132 | -0.000111072 | 0.002340483 |
| CUPP | dbCAN_3:DIAMOND | -0.000535108 | -0.000975232 | -9.50E-05 | 0.00384999 |
| dbCAN_4 | dbCAN_3:eCAMI | 0.00049709 | 6.26E-05 | 0.00093154 | 0.009700031 |
| dbCAN_4:dbCANsub | dbCAN_3:eCAMI | 0.00048837 | 5.43E-05 | 0.000922483 | 0.012279198 |
| dbCAN_3:eCAMI | dbCAN_3 | -0.000488315 | -0.000924128 | -5.25E-05 | 0.012969335 |
| CUPP | dbCAN_2:Hotpep | 0.000462335 | 2.74E-05 | 0.000897228 | 0.02533178 |
| CUPP | dbCAN_4 | -0.000462892 | -0.000901666 | -2.41E-05 | 0.027746725 |
| CUPP | dbCAN_4:dbCANsub | -0.000454172 | -0.000892613 | -1.57E-05 | 0.034221302 |
| CUPP | dbCAN_3 | -0.000454117 | -0.000894241 | -1.40E-05 | 0.035767228 |
| dbCAN_3:eCAMI | dbCAN_2:Hotpep | 0.000428138 | -2.39E-06 | 0.000858667 | 0.052928358 |
| dbCAN_3:eCAMI | dbCAN_2:DIAMOND | -0.000405783 | -0.000841253 | 2.97E-05 | 0.097131232 |
| dbCAN_4:DIAMOND | dbCAN_2:HMMER | 0.000389538 | -4.76E-05 | 0.000826646 | 0.139103353 |
| dbCAN_3:DIAMOND | dbCAN_2:HMMER | 0.00037345 | -6.37E-05 | 0.000810558 | 0.18780757 |
| CUPP | dbCAN_2:DIAMOND | -0.000371585 | -0.000811369 | 6.82E-05 | 0.202041478 |
| dbCAN_4:DIAMOND | dbCAN_3:HMMER | 0.000356622 | -8.11E-05 | 0.000794387 | 0.252614028 |
| dbCAN_3:DIAMOND | dbCAN_3:HMMER | 0.000340534 | -9.72E-05 | 0.000778299 | 0.323689973 |
| dbCAN_4:DIAMOND | dbCAN_4:HMMER | 0.000332397 | -0.000105039 | 0.000769833 | 0.362076608 |
| dbCAN_3:eCAMI | dbCAN_2 | -0.000315292 | -0.000750762 | 0.000120178 | 0.444093031 |
| dbCAN_4:HMMER | dbCAN_3:DIAMOND | -0.000316309 | -0.000753745 | 0.000121127 | 0.446296891 |
| dbCAN_4 | dbCAN_2:HMMER | 0.000301233 | -0.000134515 | 0.000736982 | 0.523368126 |
| dbCAN_4:dbCANsub | dbCAN_2:HMMER | 0.000292513 | -0.0001429 | 0.000727927 | 0.571375726 |
| dbCAN_3 | dbCAN_2:HMMER | 0.000292458 | -0.00014465 | 0.000729566 | 0.578096128 |
| CUPP | dbCAN_2 | -0.000281094 | -0.000720878 | 0.00015869 | 0.650827787 |
| dbCAN_4 | dbCAN_3:HMMER | 0.000268317 | -0.000168091 | 0.000704726 | 0.707591569 |
| dbCAN_4:DIAMOND | dbCAN_2 | 0.000270102 | -0.000169339 | 0.000709543 | 0.708001998 |
| dbCAN_4:dbCANsub | dbCAN_3:HMMER | 0.000259597 | -0.000176476 | 0.000695671 | 0.750605401 |
| dbCAN_3:HMMER | dbCAN_3 | -0.000259542 | -0.000697308 | 0.000178223 | 0.755759777 |
| dbCAN_4:HMMER | dbCAN_3:eCAMI | 0.000252997 | -0.000180449 | 0.000686443 | 0.774856463 |
| dbCAN_3:DIAMOND | dbCAN_2 | 0.000254014 | -0.000185426 | 0.000693455 | 0.786234581 |
| dbCAN_4:HMMER | dbCAN_4 | -0.000244093 | -0.00068017 | 0.000191985 | 0.821107737 |
| dbCAN_4:dbCANsub | dbCAN_4:HMMER | 0.000235373 | -0.00020037 | 0.000671115 | 0.854801522 |
| dbCAN_4:HMMER | dbCAN_3 | -0.000235318 | -0.000672754 | 0.000202118 | 0.858367221 |
| dbCAN_3:eCAMI | dbCAN_3:HMMER | -0.000228772 | -0.000662552 | 0.000205007 | 0.874629151 |
| CUPP | dbCAN_4:HMMER | -0.000218799 | -0.000656579 | 0.000218981 | 0.911522906 |
| dbCAN_2:DIAMOND | dbCAN_2:HMMER | 0.000209927 | -0.000226839 | 0.000646692 | 0.932386019 |
| dbCAN_3:eCAMI | dbCAN_2:HMMER | -0.000195856 | -0.000628972 | 0.000237259 | 0.956645408 |
| CUPP | dbCAN_3:HMMER | -0.000194575 | -0.000632684 | 0.000243535 | 0.962191648 |
| dbCAN_4 | dbCAN_2 | 0.000181798 | -0.000256291 | 0.000619887 | 0.977886368 |
| dbCAN_4:DIAMOND | dbCAN_2:DIAMOND | 0.000179611 | -0.00025983 | 0.000619052 | 0.98047187 |
| dbCAN_3:HMMER | dbCAN_2:DIAMOND | -0.000177011 | -0.000614434 | 0.000260413 | 0.982019604 |
| dbCAN_4:dbCANsub | dbCAN_2 | 0.000173078 | -0.000264677 | 0.000610833 | 0.985199109 |
| dbCAN_3 | dbCAN_2 | 0.000173023 | -0.000266418 | 0.000612464 | 0.985711447 |
| dbCAN_3:DIAMOND | dbCAN_2:DIAMOND | 0.000163523 | -0.000275918 | 0.000602964 | 0.99123264 |
| CUPP | dbCAN_2:HMMER | -0.000161659 | -0.000599111 | 0.000275794 | 0.991753512 |
| dbCAN_4:HMMER | dbCAN_2:DIAMOND | -0.000152786 | -0.00058988 | 0.000284307 | 0.995003624 |
| dbCAN_2:HMMER | dbCAN_2 | -0.000119435 | -0.000556201 | 0.00031733 | 0.999539753 |
| dbCAN_4:dbCANsub | dbCAN_4:DIAMOND | -9.70E-05 | -0.000535121 | 0.000341073 | 0.999949882 |
| dbCAN_4:DIAMOND | dbCAN_3 | 9.71E-05 | -0.000342702 | 0.000536861 | 0.999951621 |
| dbCAN_4 | dbCAN_2:DIAMOND | 9.13E-05 | -0.000346782 | 0.000529395 | 0.999974087 |
| dbCAN_2:DIAMOND | dbCAN_2 | 9.05E-05 | -0.000348609 | 0.000529591 | 0.999977094 |
| dbCAN_4:DIAMOND | dbCAN_4 | 8.83E-05 | -0.000350126 | 0.000526735 | 0.999982192 |
| dbCAN_3:HMMER | dbCAN_2 | -8.65E-05 | -0.000523943 | 0.000350904 | 0.999985415 |
| dbCAN_4:dbCANsub | dbCAN_2:DIAMOND | 8.26E-05 | -0.000355168 | 0.000520342 | 0.99999136 |
| dbCAN_3 | dbCAN_2:DIAMOND | 8.25E-05 | -0.000356909 | 0.000521973 | 0.999991783 |
| dbCAN_4:dbCANsub | dbCAN_3:DIAMOND | -8.09E-05 | -0.000519034 | 0.000357161 | 0.999993159 |
| dbCAN_3:DIAMOND | dbCAN_3 | 8.10E-05 | -0.00035879 | 0.000520773 | 0.999993396 |
| dbCAN_4 | dbCAN_3:DIAMOND | -7.22E-05 | -0.000510647 | 0.000366214 | 0.999998117 |
| dbCAN_4:HMMER | dbCAN_2 | -6.23E-05 | -0.000499388 | 0.000374799 | 0.999999639 |
| dbCAN_4:HMMER | dbCAN_2:HMMER | 5.71E-05 | -0.000377608 | 0.000491889 | 0.999999858 |
| CUPP | dbCAN_3:eCAMI | 3.42E-05 | -0.000401961 | 0.000470357 | 1 |
| dbCAN_3:HMMER | dbCAN_2:HMMER | 3.29E-05 | -0.000402164 | 0.000467996 | 1 |
| dbCAN_4:HMMER | dbCAN_3:HMMER | 2.42E-05 | -0.000411185 | 0.000459634 | 1 |
| dbCAN_4:DIAMOND | dbCAN_3:DIAMOND | 1.61E-05 | -0.000423694 | 0.000455869 | 1 |
| dbCAN_4 | dbCAN_3 | 8.78E-06 | -0.000429655 | 0.000447205 | 1 |
| dbCAN_4:dbCANsub | dbCAN_3 | 5.51E-08 | -0.000438042 | 0.000438152 | 1 |
| dbCAN_4:dbCANsub | dbCAN_4 | -8.72E-06 | -0.000445461 | 0.000428021 | 1 |

**SI Table 96: Output of Tukey HSD test, testing for statistically significant differences between the mean specificity of tools**

| Classifier 1 | Classifier 2 | Mean Difference | Lower 95% Confidence Interval | Upper 95% Confidence Interval | Adjusted P-value |
|---|---|---|---|---|---|
| dbCAN_3:DIAMOND | dbCAN_2:Hotpep | 0.000490853 | 0.00030686 | 0.000674847 | 3.75E-08 |
| dbCAN_4:DIAMOND | dbCAN_2:Hotpep | 0.000499928 | 0.000315935 | 0.000683922 | 3.75E-08 |
| dbCAN_3 | dbCAN_2:Hotpep | 0.000523606 | 0.000339613 | 0.0007076 | 3.75E-08 |
| dbCAN_4:dbCANsub | dbCAN_2:Hotpep | 0.000498765 | 0.000315493 | 0.000682037 | 3.75E-08 |
| dbCAN_2:Hotpep | dbCAN_2 | -0.000520485 | -0.000704332 | -0.000336637 | 3.75E-08 |
| CUPP | dbCAN_2:Hotpep | 0.000547876 | 0.000363736 | 0.000732017 | 3.75E-08 |
| dbCAN_4 | dbCAN_2:Hotpep | 0.000516938 | 0.000333523 | 0.000700353 | 3.75E-08 |
| dbCAN_2:Hotpep | dbCAN_2:DIAMOND | -0.000501935 | -0.000685783 | -0.000318087 | 3.75E-08 |
| dbCAN_2:Hotpep | dbCAN_2:HMMER | -0.000513068 | -0.000695916 | -0.00033022 | 3.75E-08 |
| dbCAN_3:HMMER | dbCAN_2:Hotpep | 0.000513821 | 0.000330692 | 0.000696951 | 3.75E-08 |
| dbCAN_4:HMMER | dbCAN_2:Hotpep | 0.000513049 | 0.000330061 | 0.000696038 | 3.75E-08 |
| dbCAN_3:eCAMI | dbCAN_2:Hotpep | 0.000309016 | 0.000126722 | 0.000491309 | 1.62E-06 |
| CUPP | dbCAN_3:eCAMI | 0.000238861 | 5.42E-05 | 0.000423538 | 0.001311623 |
| dbCAN_3:eCAMI | dbCAN_3 | -0.000214591 | -0.000399121 | -3.01E-05 | 0.007670145 |
| dbCAN_3:eCAMI | dbCAN_2 | -0.000211469 | -0.000395854 | -2.71E-05 | 0.009378905 |
| dbCAN_4 | dbCAN_3:eCAMI | 0.000207922 | 2.40E-05 | 0.000391876 | 0.011508599 |
| dbCAN_3:eCAMI | dbCAN_3:HMMER | -0.000204806 | -0.000388475 | -2.11E-05 | 0.013838534 |
| dbCAN_3:eCAMI | dbCAN_2:HMMER | -0.000204052 | -0.00038744 | -2.07E-05 | 0.014242711 |
| dbCAN_4:HMMER | dbCAN_3:eCAMI | 0.000204034 | 2.05E-05 | 0.000387562 | 0.014405399 |
| dbCAN_3:eCAMI | dbCAN_2:DIAMOND | -0.000192919 | -0.000377304 | -8.53E-06 | 0.03054476 |
| dbCAN_4:DIAMOND | dbCAN_3:eCAMI | 0.000190913 | 6.38E-06 | 0.000375443 | 0.034704204 |
| dbCAN_4:dbCANsub | dbCAN_3:eCAMI | 0.000189749 | 5.94E-06 | 0.00037356 | 0.035568926 |
| dbCAN_3:eCAMI | dbCAN_3:DIAMOND | -0.000181838 | -0.000366368 | 2.69E-06 | 0.058002175 |
| CUPP | dbCAN_3:DIAMOND | 5.70E-05 | -0.000129332 | 0.000243379 | 0.99858451 |
| CUPP | dbCAN_4:dbCANsub | 4.91E-05 | -0.000136531 | 0.000234754 | 0.99967225 |
| CUPP | dbCAN_4:DIAMOND | 4.79E-05 | -0.000138407 | 0.000234304 | 0.999754213 |
| CUPP | dbCAN_2:DIAMOND | 4.59E-05 | -0.00014027 | 0.000232153 | 0.99984154 |
| CUPP | dbCAN_2:HMMER | 3.48E-05 | -0.000150416 | 0.000220033 | 0.999991726 |
| CUPP | dbCAN_4:HMMER | 3.48E-05 | -0.000150536 | 0.00022019 | 0.999991747 |
| CUPP | dbCAN_3:HMMER | 3.41E-05 | -0.000151448 | 0.000219558 | 0.999993625 |
| dbCAN_3:DIAMOND | dbCAN_3 | -3.28E-05 | -0.000218964 | 0.000153457 | 0.999996052 |
| CUPP | dbCAN_4 | 3.09E-05 | -0.000154845 | 0.000216722 | 0.999997869 |
| dbCAN_3:DIAMOND | dbCAN_2 | -2.96E-05 | -0.000215698 | 0.000156435 | 0.999998716 |
| CUPP | dbCAN_2 | 2.74E-05 | -0.00015882 | 0.000213603 | 0.999999481 |
| dbCAN_4 | dbCAN_3:DIAMOND | 2.61E-05 | -0.000159554 | 0.000211723 | 0.999999693 |
| dbCAN_4:dbCANsub | dbCAN_3 | -2.48E-05 | -0.000210339 | 0.000160656 | 0.999999824 |
| CUPP | dbCAN_3 | 2.43E-05 | -0.000162085 | 0.000210626 | 0.999999872 |
| dbCAN_4:DIAMOND | dbCAN_3 | -2.37E-05 | -0.000209889 | 0.000162532 | 0.999999903 |
| dbCAN_3:DIAMOND | dbCAN_3:HMMER | -2.30E-05 | -0.000208325 | 0.000162389 | 0.999999928 |
| dbCAN_3:DIAMOND | dbCAN_2:HMMER | -2.22E-05 | -0.000207293 | 0.000162864 | 0.99999995 |
| dbCAN_4:HMMER | dbCAN_3:DIAMOND | 2.22E-05 | -0.000163021 | 0.000207414 | 0.999999951 |
| dbCAN_4:dbCANsub | dbCAN_2 | -2.17E-05 | -0.000207072 | 0.000163633 | 0.999999963 |
| dbCAN_3 | dbCAN_2:DIAMOND | 2.17E-05 | -0.000164395 | 0.000207738 | 0.999999965 |
| dbCAN_4:DIAMOND | dbCAN_2 | -2.06E-05 | -0.000206623 | 0.00016551 | 0.999999981 |
| dbCAN_2:DIAMOND | dbCAN_2 | -1.85E-05 | -0.000204472 | 0.000167372 | 0.999999994 |
| dbCAN_4:dbCANsub | dbCAN_4 | -1.82E-05 | -0.000203096 | 0.00016675 | 0.999999995 |
| dbCAN_4:DIAMOND | dbCAN_4 | -1.70E-05 | -0.000202648 | 0.000168629 | 0.999999998 |
| dbCAN_4:dbCANsub | dbCAN_3:HMMER | -1.51E-05 | -0.000199697 | 0.000169584 | 0.999999999 |
| dbCAN_4 | dbCAN_2:DIAMOND | 1.50E-05 | -0.000170491 | 0.000200497 | 1 |
| dbCAN_4:dbCANsub | dbCAN_2:HMMER | -1.43E-05 | -0.000198664 | 0.000170058 | 1 |
| dbCAN_4:dbCANsub | dbCAN_4:HMMER | -1.43E-05 | -0.000198785 | 0.000170216 | 1 |
| dbCAN_4:DIAMOND | dbCAN_3:HMMER | -1.39E-05 | -0.00019925 | 0.000171464 | 1 |
| dbCAN_4:DIAMOND | dbCAN_2:HMMER | -1.31E-05 | -0.000198218 | 0.000171939 | 1 |
| dbCAN_4:DIAMOND | dbCAN_4:HMMER | -1.31E-05 | -0.000198339 | 0.000172096 | 1 |
| dbCAN_3:HMMER | dbCAN_2:DIAMOND | 1.19E-05 | -0.000173325 | 0.000197099 | 1 |
| dbCAN_2:DIAMOND | dbCAN_2:HMMER | -1.11E-05 | -0.000196066 | 0.000173801 | 1 |
| dbCAN_4:HMMER | dbCAN_2:DIAMOND | 1.11E-05 | -0.000173958 | 0.000196187 | 1 |
| dbCAN_3:DIAMOND | dbCAN_2:DIAMOND | -1.11E-05 | -0.000197148 | 0.000174985 | 1 |
| dbCAN_4:HMMER | dbCAN_3 | -1.06E-05 | -0.000195774 | 0.000174661 | 1 |
| dbCAN_3 | dbCAN_2:HMMER | 1.05E-05 | -0.00017454 | 0.000195617 | 1 |
| dbCAN_3:HMMER | dbCAN_3 | -9.78E-06 | -0.000195142 | 0.000175572 | 1 |
| dbCAN_4:DIAMOND | dbCAN_3:DIAMOND | 9.08E-06 | -0.000177136 | 0.000195286 | 1 |
| dbCAN_2:HMMER | dbCAN_2 | -7.42E-06 | -0.00019235 | 0.000177517 | 1 |
| dbCAN_3 | dbCAN_2 | 3.12E-06 | -0.000182945 | 0.000189188 | 1 |
| dbCAN_3:HMMER | dbCAN_2 | -6.66E-06 | -0.000191875 | 0.000178549 | 1 |
| dbCAN_4 | dbCAN_2 | -3.55E-06 | -0.000189041 | 0.000181947 | 1 |
| dbCAN_4:HMMER | dbCAN_2 | -7.44E-06 | -0.000192508 | 0.000177637 | 1 |
| dbCAN_3:HMMER | dbCAN_2:HMMER | 7.54E-07 | -0.000183466 | 0.000184974 | 1 |
| dbCAN_4 | dbCAN_2:HMMER | 3.87E-06 | -0.000180633 | 0.000188373 | 1 |
| dbCAN_4:HMMER | dbCAN_2:HMMER | -1.83E-08 | -0.000184098 | 0.000184061 | 1 |
| dbCAN_4:DIAMOND | dbCAN_2:DIAMOND | -2.01E-06 | -0.000188073 | 0.00018406 | 1 |
| dbCAN_4:dbCANsub | dbCAN_2:DIAMOND | -3.17E-06 | -0.000188522 | 0.000182183 | 1 |
| dbCAN_4 | dbCAN_3 | -6.67E-06 | -0.000192307 | 0.00017897 | 1 |
| dbCAN_4 | dbCAN_3:HMMER | 3.12E-06 | -0.000181666 | 0.000187899 | 1 |
| dbCAN_4:HMMER | dbCAN_3:HMMER | -7.72E-07 | -0.000185131 | 0.000183587 | 1 |
| dbCAN_4:dbCANsub | dbCAN_3:DIAMOND | 7.91E-06 | -0.000177586 | 0.000193409 | 1 |
| dbCAN_4:HMMER | dbCAN_4 | -3.89E-06 | -0.000188531 | 0.000180754 | 1 |
| dbCAN_4:dbCANsub | dbCAN_4:DIAMOND | -1.16E-06 | -0.000186661 | 0.000184334 | 1 |

**SI Table 96: Output of Tukey HSD test, testing for statistically significant differences between the mean sensitivity of tools**

| Classifier 1 | Classifier 2 | Mean Difference | Lower 95% Confidence Interval | Upper 95% Confidence Interval | Adjusted P-value |
|---|---|---|---|---|---|
| CUPP | dbCAN_2 | 0.236597302 | 0.158015798 | 0.315178806 | 3.75E-08 |
| CUPP | dbCAN_2:DIAMOND | -0.257936142 | -0.33705143 | -0.178820854 | 3.75E-08 |
| CUPP | dbCAN_2:HMMER | -0.228527618 | -0.307109122 | -0.149946114 | 3.75E-08 |
| CUPP | dbCAN_2:Hotpep | 0.201080598 | 0.122727663 | 0.279433533 | 3.75E-08 |
| CUPP | dbCAN_3 | -0.262795674 | -0.342154389 | -0.18343696 | 3.75E-08 |
| CUPP | dbCAN_3:DIAMOND | -0.287633146 | -0.36699186 | -0.208274432 | 3.75E-08 |
| CUPP | dbCAN_3:HMMER | 0.209150282 | 0.130797347 | 0.287503217 | 3.75E-08 |
| CUPP | dbCAN_4 | -0.203690147 | -0.282271651 | -0.125108642 | 3.75E-08 |
| CUPP | dbCAN_4:dbCANsub | -0.209075723 | -0.28807125 | -0.130080197 | 3.75E-08 |
| CUPP | dbCAN_4:DIAMOND | -0.29570283 | -0.375061544 | -0.216344116 | 3.75E-08 |
| CUPP | dbCAN_4:HMMER | -0.248360377 | -0.327415624 | -0.16930513 | 3.75E-08 |
| dbCAN_2:Hotpep | dbCAN_2 | -0.202545324 | -0.281842686 | -0.123247961 | 3.75E-08 |
| dbCAN_2:Hotpep | dbCAN_2:DIAMOND | -0.215396133 | -0.294332257 | -0.136460008 | 3.75E-08 |
| dbCAN_2:Hotpep | dbCAN_2:HMMER | 0.198830614 | 0.120494951 | 0.277166277 | 3.75E-08 |
| dbCAN_3 | dbCAN_2:HMMER | 0.189254849 | 0.110979825 | 0.267529873 | 3.75E-08 |
| dbCAN_3 | dbCAN_2:Hotpep | -0.196596209 | -0.275893572 | -0.117298846 | 3.75E-08 |
| dbCAN_3:DIAMOND | dbCAN_2 | -0.171477526 | -0.250354564 | -0.092600488 | 3.75E-08 |
| dbCAN_3:DIAMOND | dbCAN_2:DIAMOND | 0.161807829 | 0.083762272 | 0.239853385 | 3.80E-08 |
| dbCAN_3:DIAMOND | dbCAN_2:HMMER | 0.171383594 | 0.09327722 | 0.249489968 | 3.75E-08 |
| dbCAN_3:DIAMOND | dbCAN_2:Hotpep | 0.176243126 | 0.097890191 | 0.254596061 | 3.75E-08 |
| dbCAN_3:eCAMI | dbCAN_2 | 0.128843584 | 0.050918693 | 0.206768476 | 3.49E-06 |
| dbCAN_3:eCAMI | dbCAN_2:DIAMOND | -0.137490681 | -0.216010227 | -0.058971136 | 5.76E-07 |
| dbCAN_3:eCAMI | dbCAN_2:HMMER | 0.124225304 | 0.045410377 | 0.203040231 | 1.40E-05 |
| dbCAN_3:eCAMI | dbCAN_3 | 0.156290605 | 0.078135892 | 0.234445317 | 4.05E-08 |
| dbCAN_3:eCAMI | dbCAN_3:DIAMOND | -0.149970196 | -0.228184903 | -0.071755488 | 5.54E-08 |
| dbCAN_3:eCAMI | dbCAN_3:HMMER | -0.143439796 | -0.221959342 | -0.06492025 | 1.56E-07 |
| dbCAN_3:HMMER | dbCAN_2:Hotpep | 0.122523175 | 0.044538112 | 0.200508239 | 1.53E-05 |
| dbCAN_4 | dbCAN_2:HMMER | -0.112371998 | -0.190467033 | -0.034276963 | 0.000142486 |
| dbCAN_4 | dbCAN_2:Hotpep | 0.11615562 | 0.037340693 | 0.194970547 | 8.07E-05 |
| dbCAN_4 | dbCAN_3:eCAMI | -0.115992775 | -0.194283571 | -0.03770198 | 7.09E-05 |
| dbCAN_4:dbCANsub | dbCAN_2:Hotpep | 0.099106621 | 0.01987104 | 0.178342202 | 0.002392306 |
| dbCAN_4:dbCANsub | dbCAN_3:eCAMI | -0.110043661 | -0.188334456 | -0.031752865 | 0.000242274 |
| dbCAN_4:DIAMOND | dbCAN_2 | -0.086552548 | -0.16496796 | -0.008137136 | 0.015837066 |
| dbCAN_4:DIAMOND | dbCAN_2:DIAMOND | 0.086458616 | 0.0078888 | 0.165028432 | 0.016479631 |
| dbCAN_4:DIAMOND | dbCAN_2:HMMER | 0.091036937 | 0.011801356 | 0.170272518 | 0.009142247 |
| dbCAN_4:DIAMOND | dbCAN_2:Hotpep | 0.093157506 | 0.013921925 | 0.172393087 | 0.006514468 |
| dbCAN_4:DIAMOND | dbCAN_3:eCAMI | 0.091318149 | 0.012503221 | 0.170133076 | 0.008094702 |
| dbCAN_4:DIAMOND | dbCAN_3:HMMER | 0.086627106 | 0.007693598 | 0.165560615 | 0.01705926 |
| dbCAN_4:DIAMOND | dbCAN_4:HMMER | -0.084924978 | -0.162790015 | -0.00705994 | 0.018472699 |
| dbCAN_4:HMMER | dbCAN_2:Hotpep | 0.080306697 | 0.001432637 | 0.159180757 | 0.04137017 |
| dbCAN_4:HMMER | dbCAN_3:eCAMI | 0.085087822 | 0.005852241 | 0.164323403 | 0.022425113 |
| dbCAN_3:DIAMOND | dbCAN_3:HMMER | 0.078557423 | -0.000376086 | 0.157490931 | 0.052502805 |
| dbCAN_4:dbCANsub | dbCAN_2:HMMER | 0.076882851 | -0.001626507 | 0.155392209 | 0.061669139 |
| dbCAN_4:HMMER | dbCAN_3:DIAMOND | -0.072237013 | -0.151111073 | 0.006637047 | 0.112231378 |
| dbCAN_3 | dbCAN_2 | 0.066199466 | -0.013036116 | 0.145435047 | 0.217022824 |
| dbCAN_4 | dbCAN_2 | 0.061339933 | -0.017651842 | 0.140331708 | 0.32650116 |
| dbCAN_3 | dbCAN_2:DIAMOND | 0.060250351 | -0.01898523 | 0.139485932 | 0.360941313 |
| CUPP | dbCAN_3:eCAMI | -0.059105528 | -0.137749327 | 0.019538272 | 0.380491819 |
| dbCAN_4 | dbCAN_2:DIAMOND | 0.055390818 | -0.023600957 | 0.134382594 | 0.499061016 |
| dbCAN_3:HMMER | dbCAN_3 | -0.053719951 | -0.13265346 | 0.025213558 | 0.549777634 |
| dbCAN_4:dbCANsub | dbCAN_2 | 0.051764168 | -0.027167473 | 0.130695808 | 0.610609494 |
| dbCAN_4 | dbCAN_3:HMMER | 0.048860419 | -0.029828348 | 0.127549185 | 0.693645054 |
| dbCAN_4:HMMER | dbCAN_3 | -0.047399542 | -0.126273602 | 0.031474518 | 0.738441663 |
| dbCAN_4:dbCANsub | dbCAN_4:DIAMOND | -0.047342453 | -0.126335729 | 0.031650823 | 0.74197697 |
| dbCAN_4:dbCANsub | dbCAN_2:DIAMOND | 0.045815053 | -0.033116587 | 0.124746694 | 0.781433213 |
| dbCAN_4:HMMER | dbCAN_2:HMMER | 0.043918607 | -0.0344708 | 0.122308013 | 0.820168343 |
| dbCAN_4:HMMER | dbCAN_4 | -0.042540009 | -0.121169143 | 0.036089124 | 0.853411054 |
| dbCAN_4:dbCANsub | dbCAN_3:HMMER | 0.039284653 | -0.039343747 | 0.117913054 | 0.911722737 |
| dbCAN_4:dbCANsub | dbCAN_3:DIAMOND | -0.039272769 | -0.118266045 | 0.039720507 | 0.914609891 |
| dbCAN_3:HMMER | dbCAN_2:HMMER | 0.037598197 | -0.040851025 | 0.11604742 | 0.933735477 |
| dbCAN_4:DIAMOND | dbCAN_4 | 0.037766688 | -0.041286676 | 0.116820052 | 0.935220099 |
| dbCAN_4:dbCANsub | dbCAN_4:HMMER | 0.032964244 | -0.045604476 | 0.111532965 | 0.975830263 |
| dbCAN_4:DIAMOND | dbCAN_3 | 0.032907155 | -0.046389825 | 0.112204136 | 0.977883596 |
| dbCAN_2:DIAMOND | dbCAN_2:HMMER | 0.031067798 | -0.047685355 | 0.10982095 | 0.985475581 |
| dbCAN_4 | dbCAN_3:DIAMOND | -0.029697004 | -0.108750368 | 0.04935636 | 0.990471625 |
| dbCAN_3:eCAMI | dbCAN_2:Hotpep | -0.027447021 | -0.105075779 | 0.050181738 | 0.994454382 |
| dbCAN_2:HMMER | dbCAN_2 | -0.025118683 | -0.103871835 | 0.053634469 | 0.997884324 |
| dbCAN_3:DIAMOND | dbCAN_3 | 0.024837471 | -0.054459509 | 0.104134452 | 0.998223628 |
| dbCAN_4:HMMER | dbCAN_2 | 0.018799924 | -0.060012408 | 0.097612255 | 0.999889014 |
| dbCAN_4:dbCANsub | dbCAN_3 | -0.014435298 | -0.093428573 | 0.064557978 | 0.999993943 |
| dbCAN_4:HMMER | dbCAN_2:DIAMOND | 0.012850809 | -0.065961523 | 0.09166314 | 0.999998321 |
| dbCAN_3:HMMER | dbCAN_2 | 0.012479514 | -0.066392312 | 0.091351341 | 0.999998807 |
| dbCAN_4:dbCANsub | dbCAN_4 | -0.009575765 | -0.088324485 | 0.069172955 | 0.999999942 |
| dbCAN_4:DIAMOND | dbCAN_3:DIAMOND | 0.008069684 | -0.071227296 | 0.087366664 | 0.999999993 |
| dbCAN_3:HMMER | dbCAN_2:DIAMOND | 0.0065304 | -0.072341427 | 0.085402226 | 0.999999999 |
| dbCAN_4:HMMER | dbCAN_3:HMMER | 0.006320409 | -0.072188221 | 0.08482904 | 1 |
| dbCAN_2:DIAMOND | dbCAN_2 | 0.005949115 | -0.07322502 | 0.085123249 | 1 |
| dbCAN_4 | dbCAN_3 | -0.004859533 | -0.083912896 | 0.074193831 | 1 |

**SI Table 97: Output of Tukey HSD test, testing for statistically significant differences between the mean precision of tools**

| Classifier 1 | Classifier 2 | Mean Difference | Lower 95% Confidence Interval | Upper 95% Confidence Interval | Adjusted P-value |
|---|---|---|---|---|---|
| CUPP | dbCAN_2 | -0.196643575 | -0.278692064 | -0.114595086 | 3.75E-08 |
| CUPP | dbCAN_2:DIAMOND | -0.194297426 | -0.276345915 | -0.112248937 | 3.75E-08 |
| CUPP | dbCAN_2:HMMER | -0.149903849 | -0.23151743 | -0.068290267 | 1.34E-07 |
| CUPP | dbCAN_3 | -0.234771483 | -0.316883452 | -0.152659514 | 3.75E-08 |
| CUPP | dbCAN_3:DIAMOND | -0.234000228 | -0.316112197 | -0.151888259 | 3.75E-08 |
| CUPP | dbCAN_3:HMMER | -0.183559649 | -0.26529583 | -0.101823469 | 3.75E-08 |
| CUPP | dbCAN_4 | -0.211296307 | -0.293156404 | -0.12943621 | 3.75E-08 |
| CUPP | dbCAN_4:dbCANsub | -0.201768038 | -0.283566011 | -0.119970066 | 3.75E-08 |
| CUPP | dbCAN_4:DIAMOND | -0.243785722 | -0.325897691 | -0.161673753 | 3.75E-08 |
| CUPP | dbCAN_4:HMMER | -0.190187127 | -0.271861845 | -0.108512409 | 3.75E-08 |
| dbCAN_2:Hotpep | dbCAN_2 | -0.152793934 | -0.233800934 | -0.071786934 | 7.21E-08 |
| dbCAN_2:Hotpep | dbCAN_2:DIAMOND | -0.150447785 | -0.231454785 | -0.069440785 | 1.01E-07 |
| dbCAN_2:Hotpep | dbCAN_2:HMMER | -0.106054208 | -0.186620679 | -0.025487737 | 0.000939967 |
| dbCAN_3 | dbCAN_2:HMMER | 0.084867634 | 0.003318318 | 0.16641695 | 0.03247609 |
| dbCAN_3 | dbCAN_2:Hotpep | 0.190921842 | 0.109850547 | 0.271993138 | 3.75E-08 |
| dbCAN_3:DIAMOND | dbCAN_2:HMMER | 0.084096379 | 0.002547063 | 0.165645695 | 0.035983838 |
| dbCAN_3:DIAMOND | dbCAN_2:Hotpep | 0.190150587 | 0.109079292 | 0.271221883 | 3.75E-08 |
| dbCAN_3:eCAMI | dbCAN_2 | -0.138222199 | -0.219465886 | -0.056978513 | 1.44E-06 |
| dbCAN_3:eCAMI | dbCAN_2:DIAMOND | -0.13587605 | -0.217119736 | -0.054632364 | 2.46E-06 |
| dbCAN_3:eCAMI | dbCAN_2:HMMER | -0.091482473 | -0.172286921 | -0.010678025 | 0.011251732 |
| dbCAN_3:eCAMI | dbCAN_3 | -0.176350107 | -0.257657902 | -0.095042313 | 3.75E-08 |
| dbCAN_3:eCAMI | dbCAN_3:DIAMOND | -0.175578852 | -0.256886647 | -0.094271057 | 3.76E-08 |
| dbCAN_3:eCAMI | dbCAN_3:HMMER | -0.125138274 | -0.206066546 | -0.044210001 | 2.37E-05 |
| dbCAN_3:HMMER | dbCAN_2:Hotpep | 0.139710009 | 0.059019348 | 0.22040067 | 8.26E-07 |
| dbCAN_4 | dbCAN_2:Hotpep | 0.167446667 | 0.086630486 | 0.248262847 | 3.81E-08 |
| dbCAN_4 | dbCAN_3:eCAMI | 0.152874932 | 0.071821508 | 0.233928355 | 7.22E-08 |
| dbCAN_4:dbCANsub | dbCAN_2:Hotpep | 0.157918398 | 0.077165145 | 0.238671651 | 4.53E-08 |
| dbCAN_4:dbCANsub | dbCAN_3:eCAMI | 0.143346663 | 0.062355982 | 0.224337344 | 4.09E-07 |
| dbCAN_4:DIAMOND | dbCAN_2:HMMER | 0.093881873 | 0.012332557 | 0.175431189 | 0.008884575 |
| dbCAN_4:DIAMOND | dbCAN_2:Hotpep | 0.199936081 | 0.118864786 | 0.281007377 | 3.75E-08 |
| dbCAN_4:DIAMOND | dbCAN_3:eCAMI | 0.185364346 | 0.104056551 | 0.266672141 | 3.75E-08 |
| dbCAN_4:HMMER | dbCAN_2:Hotpep | 0.146337486 | 0.065709085 | 0.226965888 | 1.88E-07 |
| dbCAN_4:HMMER | dbCAN_3:eCAMI | 0.131765751 | 0.050899555 | 0.212631947 | 5.41E-06 |
| dbCAN_4 | dbCAN_2:HMMER | 0.061392459 | -0.019903242 | 0.14268816 | 0.372436712 |
| dbCAN_4:DIAMOND | dbCAN_3:HMMER | 0.060226072 | -0.021445939 | 0.141898084 | 0.412708471 |
| CUPP | dbCAN_3:eCAMI | -0.058421376 | -0.139793627 | 0.022950875 | 0.458575531 |
| dbCAN_4:DIAMOND | dbCAN_4:HMMER | 0.053598595 | -0.028011905 | 0.135209095 | 0.608289145 |
| dbCAN_4:dbCANsub | dbCAN_2:HMMER | 0.05186419 | -0.029368955 | 0.133097335 | 0.652507185 |
| dbCAN_3:HMMER | dbCAN_3 | -0.051211833 | -0.132883845 | 0.030460178 | 0.679472924 |
| dbCAN_3:DIAMOND | dbCAN_3:HMMER | 0.050440578 | -0.031231433 | 0.13211259 | 0.701290696 |
| dbCAN_4:DIAMOND | dbCAN_2:DIAMOND | 0.049488296 | -0.032496268 | 0.13147286 | 0.732592322 |
| dbCAN_4:DIAMOND | dbCAN_2 | 0.047142147 | -0.034842417 | 0.129126711 | 0.792219836 |
| dbCAN_2:HMMER | dbCAN_2 | -0.046739726 | -0.128225124 | 0.034745672 | 0.79499108 |
| dbCAN_4:HMMER | dbCAN_3 | -0.044584356 | -0.126194856 | 0.037026144 | 0.844655625 |
| dbCAN_2:DIAMOND | dbCAN_2:HMMER | 0.044393577 | -0.037091821 | 0.125878975 | 0.847173989 |
| CUPP | dbCAN_2:Hotpep | -0.043849641 | -0.124985581 | 0.037286299 | 0.854344975 |
| dbCAN_4:HMMER | dbCAN_3:DIAMOND | -0.043813101 | -0.125423601 | 0.037797399 | 0.860104274 |
| dbCAN_4:dbCANsub | dbCAN_4:DIAMOND | -0.042017683 | -0.123751535 | 0.039716169 | 0.893445551 |
| dbCAN_4:HMMER | dbCAN_2:HMMER | 0.040283278 | -0.040825754 | 0.12139231 | 0.915200999 |
| dbCAN_3 | dbCAN_2:DIAMOND | 0.040474057 | -0.041510507 | 0.122458621 | 0.918596202 |
| dbCAN_3:DIAMOND | dbCAN_2:DIAMOND | 0.039702802 | -0.042281762 | 0.121687366 | 0.92871977 |
| dbCAN_3 | dbCAN_2 | 0.038127908 | -0.043856656 | 0.120112472 | 0.946624069 |
| dbCAN_3:DIAMOND | dbCAN_2 | 0.037356653 | -0.044627911 | 0.119341217 | 0.954103258 |
| dbCAN_3:HMMER | dbCAN_2:HMMER | 0.033655801 | -0.047515122 | 0.114826724 | 0.978038203 |
| dbCAN_4:dbCANsub | dbCAN_3 | -0.033003444 | -0.114737296 | 0.048730408 | 0.982340085 |
| dbCAN_4:DIAMOND | dbCAN_4 | 0.032489415 | -0.04930661 | 0.114285439 | 0.984612927 |
| dbCAN_4:dbCANsub | dbCAN_3:DIAMOND | -0.032232189 | -0.113966041 | 0.049501663 | 0.985519318 |
| dbCAN_4 | dbCAN_3:HMMER | 0.027736658 | -0.053682121 | 0.109155437 | 0.99605944 |
| dbCAN_4 | dbCAN_3 | -0.023475176 | -0.1052712 | 0.058320849 | 0.999249606 |
| dbCAN_4 | dbCAN_3:DIAMOND | -0.022703921 | -0.104499945 | 0.059092104 | 0.999464266 |
| dbCAN_4:HMMER | dbCAN_4 | -0.021109181 | -0.102466257 | 0.060247896 | 0.99973183 |
| dbCAN_4:dbCANsub | dbCAN_3:HMMER | 0.018208389 | -0.063147928 | 0.099564707 | 0.999943859 |
| dbCAN_4 | dbCAN_2:DIAMOND | 0.016998881 | -0.064733418 | 0.098731181 | 0.999974676 |
| dbCAN_3:eCAMI | dbCAN_2:Hotpep | 0.014571735 | -0.065750259 | 0.09489373 | 0.999994416 |
| dbCAN_4 | dbCAN_2 | 0.014652732 | -0.067079567 | 0.096385032 | 0.999995111 |
| dbCAN_3:HMMER | dbCAN_2 | -0.013083925 | -0.094692115 | 0.068524264 | 0.999998614 |
| dbCAN_4:dbCANsub | dbCAN_4:HMMER | 0.011580912 | -0.069713656 | 0.092875479 | 0.999999641 |
| dbCAN_3:HMMER | dbCAN_2:DIAMOND | -0.010737776 | -0.092345966 | 0.070870413 | 0.999999856 |
| dbCAN_4:DIAMOND | dbCAN_3:DIAMOND | 0.009785494 | -0.072262599 | 0.091833587 | 0.999999954 |
| dbCAN_4:dbCANsub | dbCAN_4 | -0.009528269 | -0.09100908 | 0.071952543 | 0.999999963 |
| dbCAN_4:DIAMOND | dbCAN_3 | 0.009014239 | -0.073033854 | 0.091062332 | 0.999999982 |
| dbCAN_4:dbCANsub | dbCAN_2:DIAMOND | 0.007470613 | -0.074199466 | 0.089140691 | 0.999999998 |
| dbCAN_4:HMMER | dbCAN_3:HMMER | 0.006627477 | -0.074604915 | 0.08785987 | 0.999999999 |
| dbCAN_4:HMMER | dbCAN_2 | -0.006456448 | -0.088003078 | 0.075090182 | 1 |
| dbCAN_4:dbCANsub | dbCAN_2 | 0.005124464 | -0.076545615 | 0.086794542 | 1 |
| dbCAN_4:HMMER | dbCAN_2:DIAMOND | -0.004110299 | -0.085656929 | 0.077436331 | 1 |
| dbCAN_2:DIAMOND | dbCAN_2 | -0.002346149 | -0.084267134 | 0.079574836 | 1 |
| dbCAN_3:DIAMOND | dbCAN_3 | -0.000771255 | -0.082819348 | 0.081276838 | 1 |

## 12 Performance of CAZy family classification per CAZy class

**SI table 98: The number and proportion of test sets with an F1-score greater than 0.9 for CAZyme family classification (overleaf)**

| | GH (%) | | | GT (%) | | | PL (%) | | | CE (%) | | | AA (%) | | | CBM (%) | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **Prediction Tool** | **1** | **≥0.9** | **≤0.75** | **1** | **≥0.9** | **≤0.75** | **1** | **≥0.9** | **≤0.75** | **1** | **≥0.9** | **≤0.75** | **1** | **≥0.9** | **≤0.75** | **1** | **≥0.9** | **≤0.75** |
| **dbCAN_2** | 49.265 | 80.147 | 11.029 | 37.333 | 73.333 | 14.667 | 60.870 | 69.565 | 17.391 | 16.667 | 44.444 | 38.889 | 18.750 | 50.000 | 25.000 | 47.273 | 56.364 | 23.636 |
| **dbCAN_2:HMMER** | 46.429 | 75.714 | 12.857 | 39.474 | 65.789 | 18.421 | 69.565 | 73.913 | 21.739 | 21.053 | 42.105 | 42.105 | 37.500 | 56.250 | 25.000 | 44.643 | 51.786 | 30.357 |
| **dbCAN_2:DIAMOND** | 45.588 | 80.147 | 11.765 | 40.000 | 80.000 | 14.667 | 65.217 | 78.261 | 13.043 | 16.667 | 38.889 | 33.333 | 12.500 | 37.500 | 18.750 | 43.636 | 52.727 | 21.818 |
| **dbCAN_2:Hotpep** | 32.117 | 64.234 | 21.168 | 33.333 | 62.667 | 26.667 | 37.500 | 50.000 | 33.333 | 5.882 | 23.529 | 47.059 | 18.750 | 50.000 | 37.500 | 4.348 | 5.797 | 88.406 |
| **dbCAN_3** | 57.778 | 88.889 | 5.185 | 48.684 | 81.579 | 11.842 | 78.261 | 86.957 | 8.696 | 29.412 | 58.824 | 23.529 | 37.500 | 62.500 | 6.250 | 52.727 | 65.455 | 16.364 |
| **dbCAN_3:HMMER** | 48.921 | 80.576 | 10.072 | 42.105 | 68.421 | 15.789 | 69.565 | 78.261 | 13.043 | 27.778 | 50.000 | 33.333 | 37.500 | 62.500 | 18.750 | 46.429 | 53.571 | 28.571 |
| **dbCAN_3:DIAMOND** | 60.000 | 89.630 | 3.704 | 55.263 | 86.842 | 9.211 | 73.913 | 86.957 | 13.043 | 29.412 | 58.824 | 29.412 | 37.500 | 56.250 | 12.500 | 52.727 | 76.364 | 14.545 |
| **dbCAN_3:eCAMI** | 21.168 | 59.124 | 25.547 | 31.579 | 64.474 | 25.000 | 34.783 | 52.174 | 39.130 | 11.765 | 29.412 | 47.059 | 12.500 | 25.000 | 31.250 | 7.692 | 12.308 | 75.385 |
| **dbCAN_4** | 56.934 | 87.591 | 6.569 | 47.368 | 80.263 | 11.842 | 69.565 | 78.261 | 13.043 | 33.333 | 61.111 | 22.222 | 37.500 | 62.500 | 18.750 | 55.357 | 73.214 | 19.643 |
| **dbCAN_4:HMMER** | 49.640 | 81.295 | 9.353 | 41.558 | 67.532 | 16.883 | 69.565 | 78.261 | 13.043 | 27.778 | 55.556 | 22.222 | 37.500 | 62.500 | 18.750 | 46.429 | 55.357 | 26.786 |
| **dbCAN_4:DIAMOND** | 60.741 | 90.370 | 3.704 | 52.632 | 85.526 | 9.211 | 78.261 | 86.957 | 13.043 | 47.059 | 70.588 | 17.647 | 37.500 | 56.250 | 12.500 | 52.727 | 80.000 | 14.545 |
| **dbCAN_4:dbCAN-sub** | 52.899 | 85.507 | 8.696 | 46.053 | 80.263 | 11.842 | 69.565 | 78.261 | 13.043 | 33.333 | 61.111 | 22.222 | 37.500 | 62.500 | 18.750 | 46.429 | 69.643 | 25.000 |
| **CUPP** | 41.481 | 70.370 | 20.000 | 33.333 | 69.333 | 22.667 | 52.174 | 60.870 | 34.783 | 35.294 | 58.824 | 35.294 | 25.000 | 56.250 | 31.250 | 0.000 | 0.000 | 100.000 |
| **Mean** | 47.920 | 79.507 | 11.511 | 42.209 | 74.310 | 16.054 | 63.754 | 73.746 | 18.952 | 25.802 | 50.247 | 31.871 | 29.808 | 53.846 | 21.154 | 38.494 | 50.199 | 37.312 |
| **Standard Deviation** | 10.862 | 9.453 | 6.595 | 7.176 | 8.087 | 5.502 | 13.491 | 11.891 | 9.705 | 10.643 | 13.231 | 9.562 | 10.153 | 10.836 | 8.327 | 19.253 | 25.907 | 28.544 |

## 12.1 Specificity

### SI Figure 34: Confidence intervals of the specificity of the binary classification

The 95% confidence intervals (CI) of the specificity for each CAZy family was calculated, when pooling all test sets.

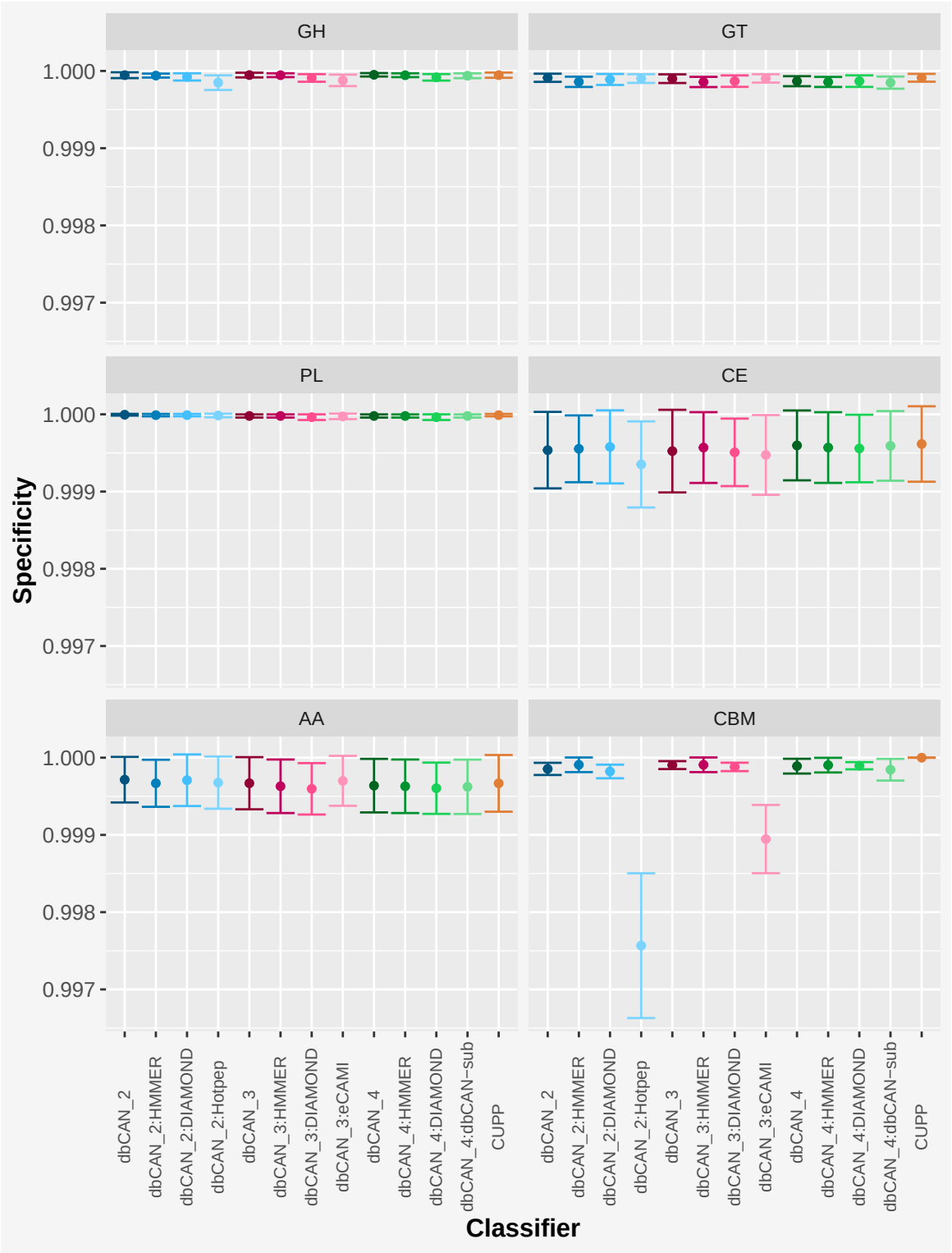



Figure 34: 95% confidence intervals of the specificity scores calculated for each CAZy family when all test sets are pooled. Where a tool is incorporated into a CAZyme classifier, the tool name is prefixed with the respective CAZyme classifier.

## SI Figure 35: Specificity of the binary classification of each CAZy family

The specificity for each CAZy family was calculated, when pooling all test sets.

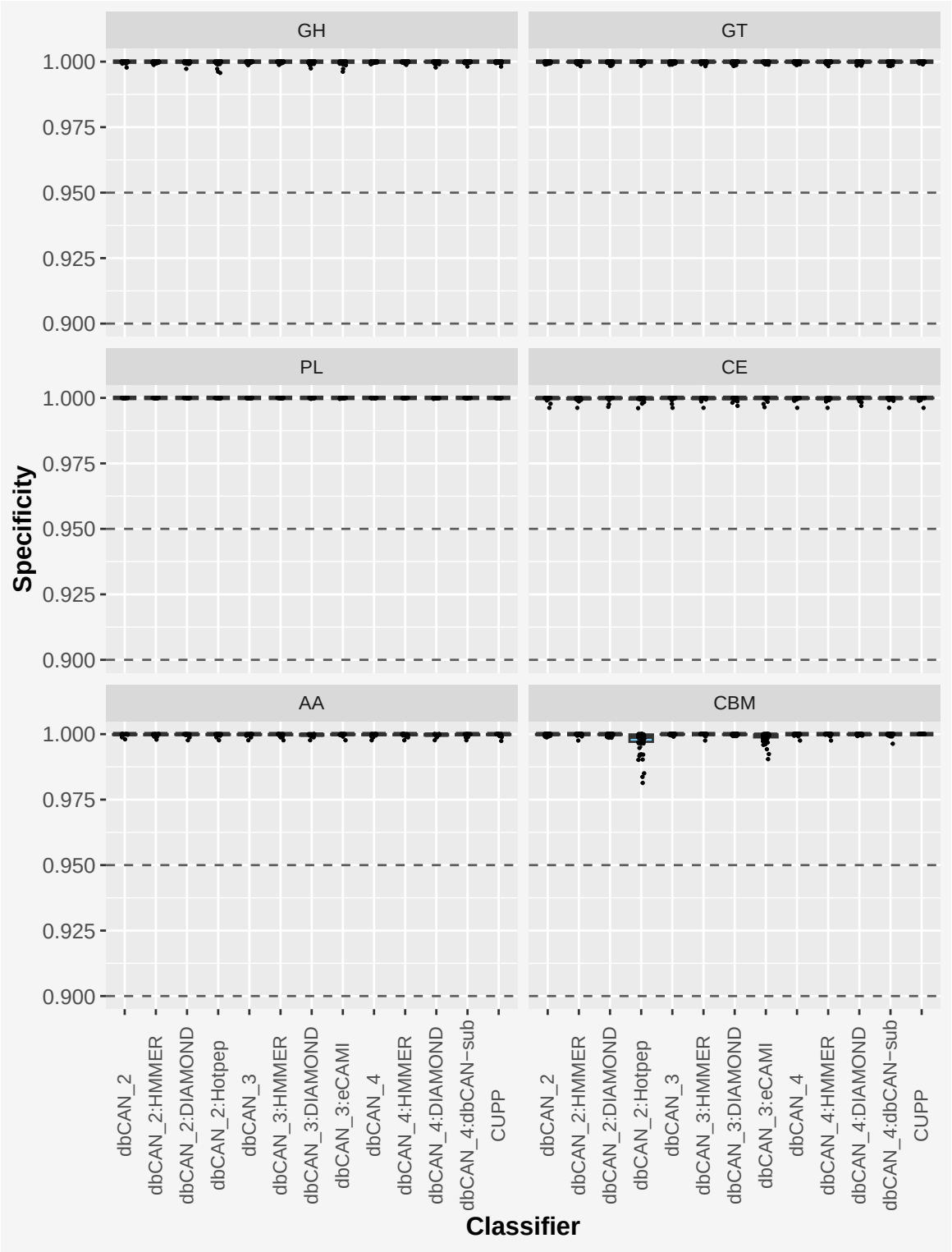



Figure 35: Specificity of each CAZy family, when pooling all test sets. One-dimensional scatter plot overlaying a box and whisker plot, where each point in the scatter plot represents a CAZy family. Where a tool is incorporated into a CAZyme classifier, the tool name is prefixed with the respective CAZyme classifier.

## 12.2 Sensitivity

### SI Figure 36: Confidence intervals of the sensitivity of the binary classification

The 95% confidence intervals (CI) of the sensitivity for each CAZy family was calculated, when pooling all test sets.

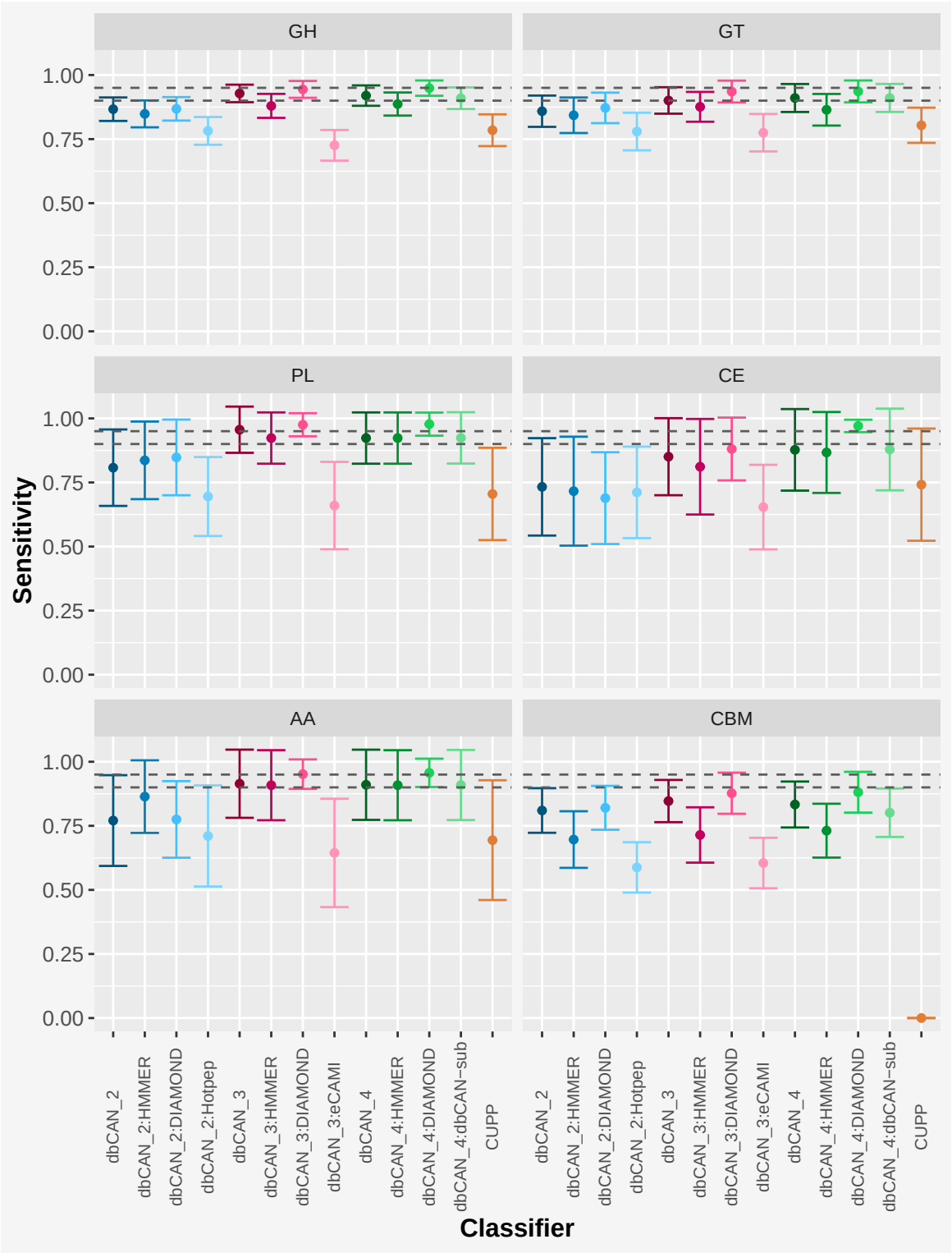



Figure 36: 95% confidence intervals of the sensitivity scores calculated for each CAZy family when all test sets are pooled. Where a tool is incorporated into a CAZyme classifier, the tool name is prefixed with the respective CAZyme classifier.

## SI Figure 37: Sensitivity of the binary classification of each CAZy family

The sensitivity for each CAZy family was calculated, when pooling all test sets.

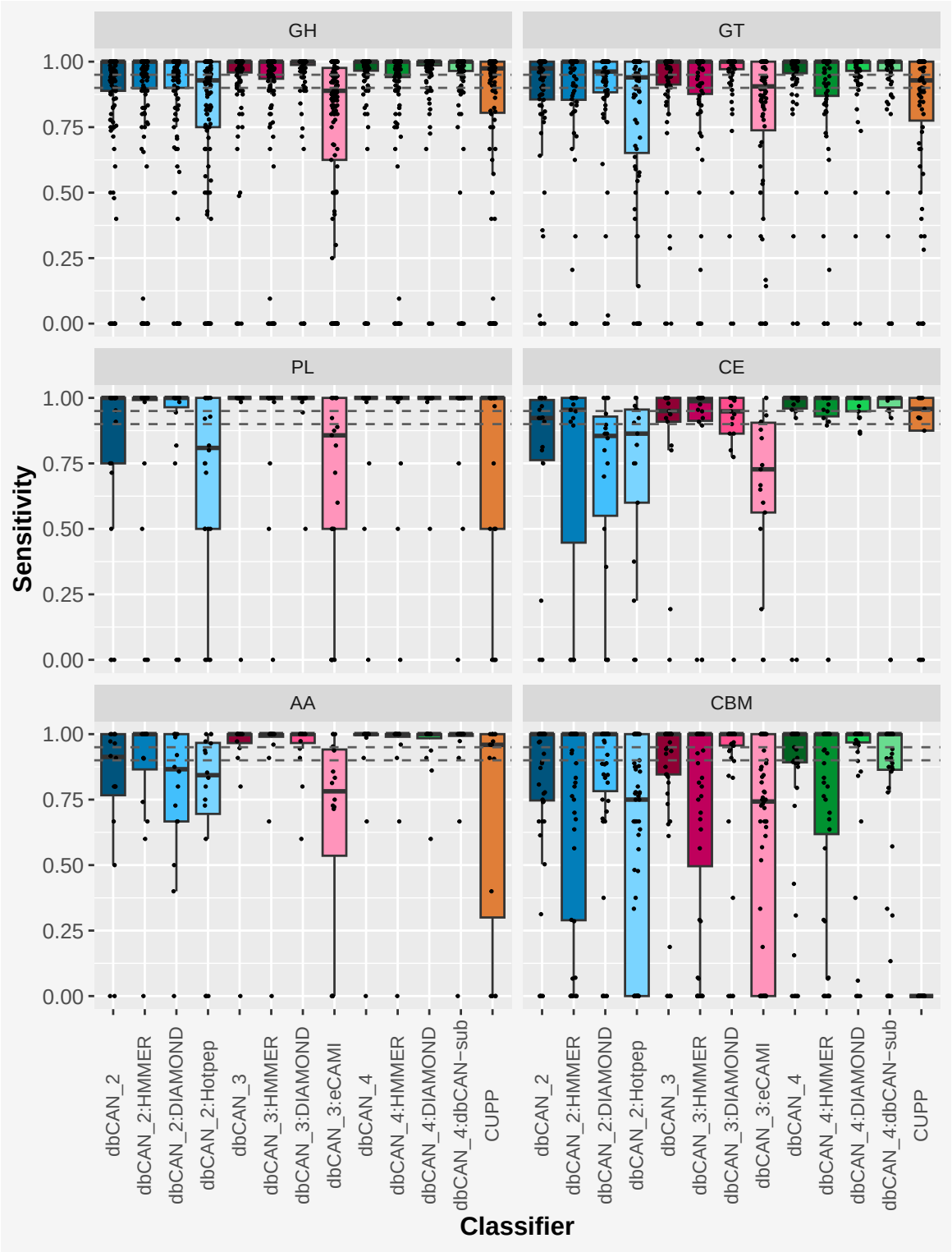



Figure 37: Sensitivity of each CAZy family, when pooling all test sets. One-dimensional scatter plot overlaying a box and whisker plot, where each point in the scatter plot represents a CAZy family. Where a tool is incorporated into a CAZyme classifier, the tool name is prefixed with the respective CAZyme classifier.

## 12.3 Precision

### SI Figure 38: Confidence intervals of the precision of the binary classification

The 95% confidence intervals (CI) of the precision for each CAZy family was calculated, when pooling all test sets.

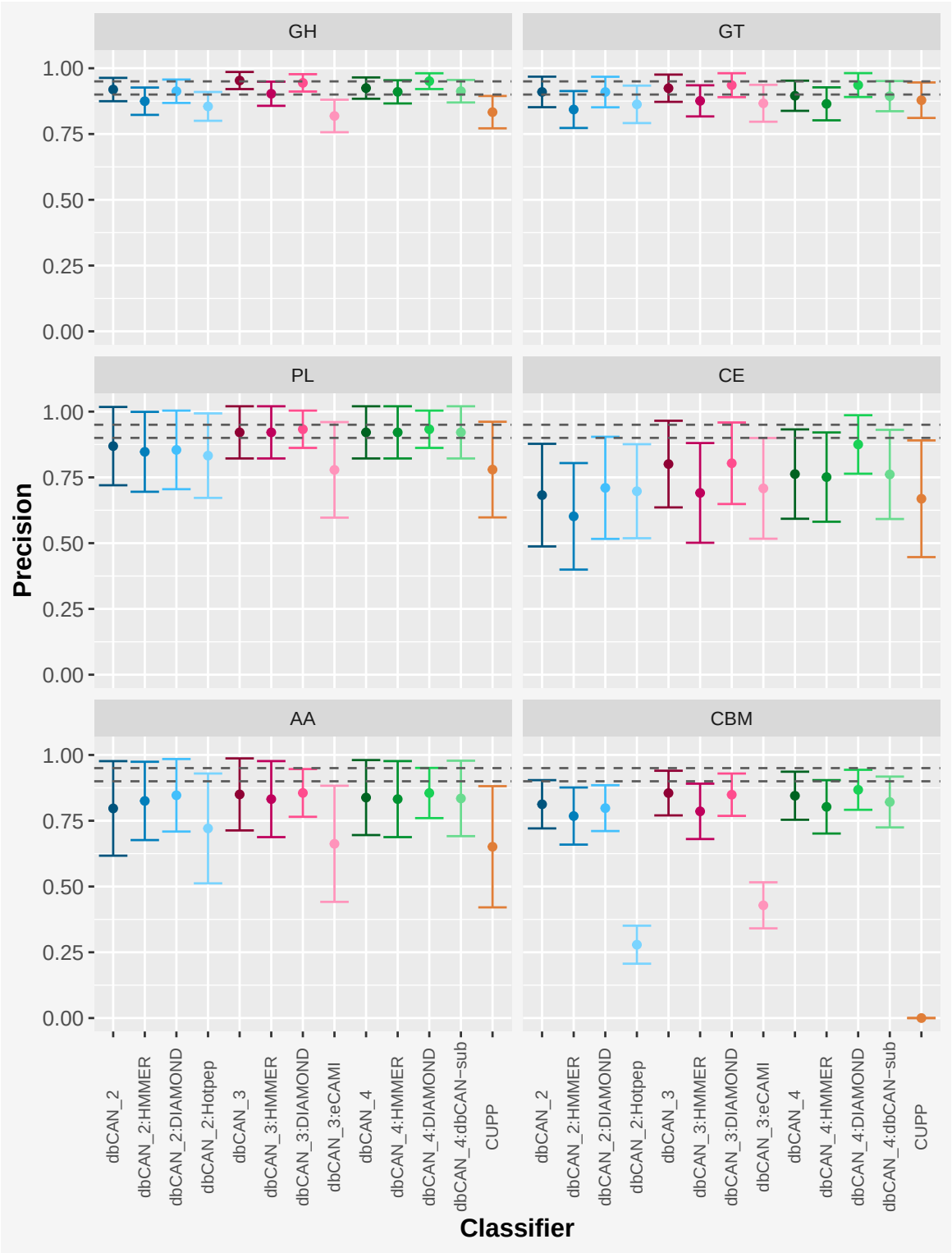



Figure 38: 95% confidence intervals of the precision scores calculated for each CAZy family when all test sets are pooled. Where a tool is incorporated into a CAZyme classifier, the tool name is prefixed with the respective CAZyme classifier.

## SI Figure 39: Precision of the binary classification of each CAZy family

The precision for each CAZy family was calculated, when pooling all test sets.

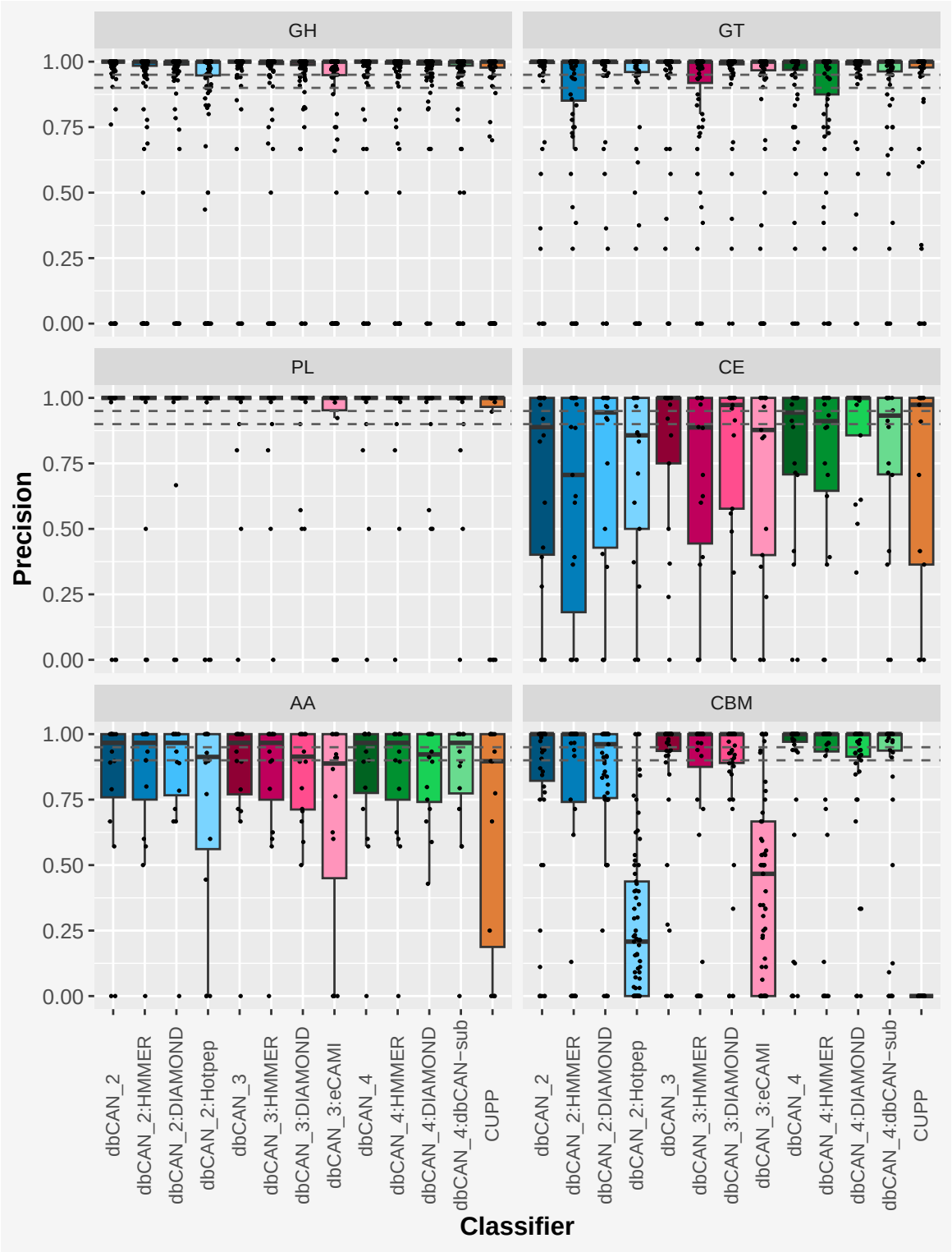



Figure 39: Precision of each CAZy family, when pooling all test sets. One-dimensional scatter plot overlaying a box and whisker plot, where each point in the scatter plot represents a CAZy family. Where a tool is incorporated into a CAZyme classifier, the tool name is prefixed with the respective CAZyme classifier.

## 12.4 F1-scpre

### SI Figure 40: Confidence intervals of the F1-score of the binary classification

The 95% confidence intervals (CI) of the F1-score for each CAZy family was calculated, when pooling all test sets.

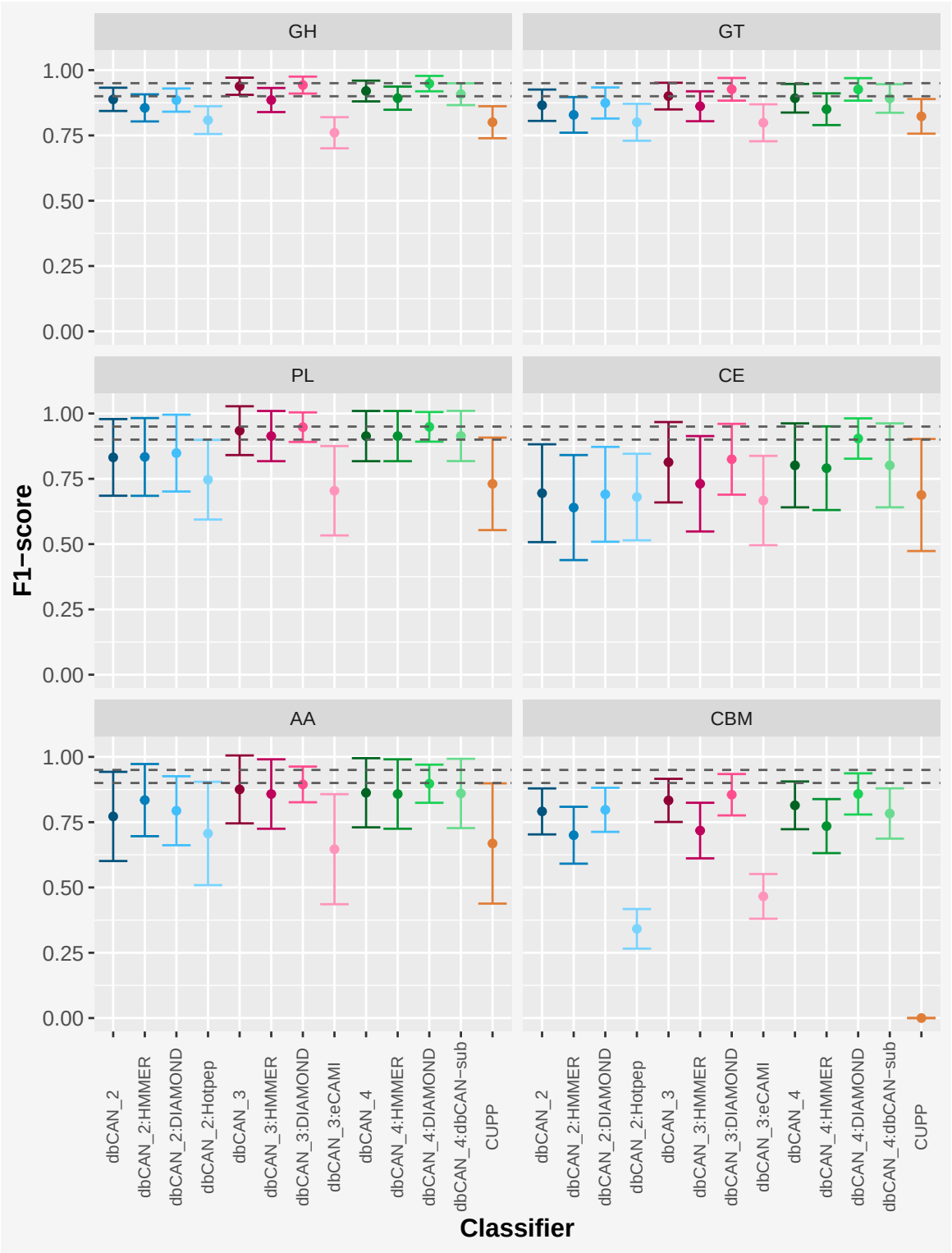



Figure 40: 95% confidence intervals of the F1-score scores calculated for each CAZy family when all test sets are pooled. Where a tool is incorporated into a CAZyme classifier, the tool name is prefixed with the respective CAZyme classifier.

## SI Figure 41: F1-score of the binary classification of each CAZy family

The F1-score for each CAZy family was calculated, when pooling all test sets.

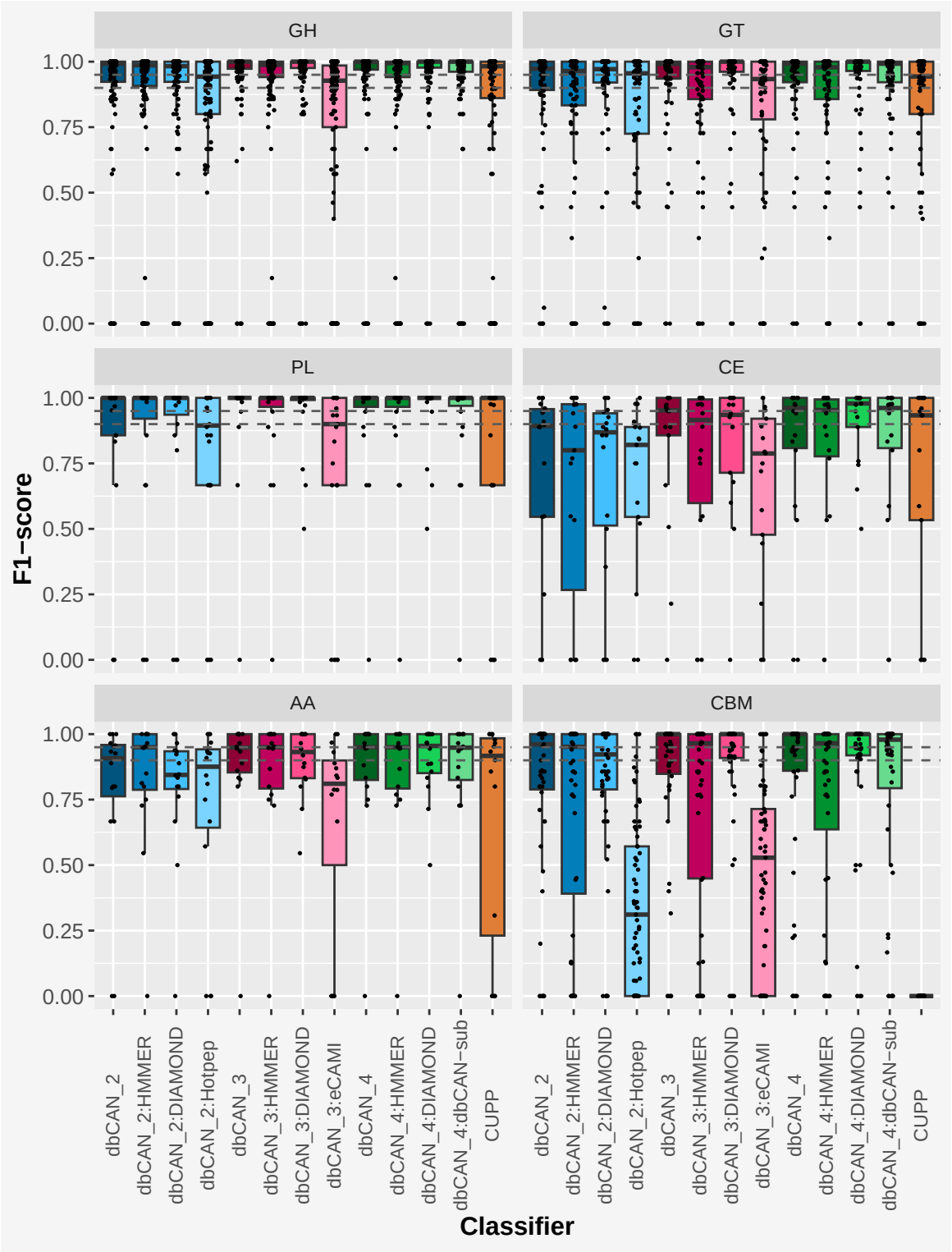



Figure 41: F1-score of each CAZy family, when pooling all test sets. One-dimensional scatter plot overlaying a box and whisker plot, where each point in the scatter plot represents a CAZy family. Where a tool is incorporated into a CAZyme classifier, the tool name is prefixed with the respective CAZyme classifier.

## 12.5 Accuracy

### SI Figure 42: Confidence intervals of the accuracy of the binary classification

The 95% confidence intervals (CI) of the accuracy for each CAZy family was calculated, when pooling all test sets.

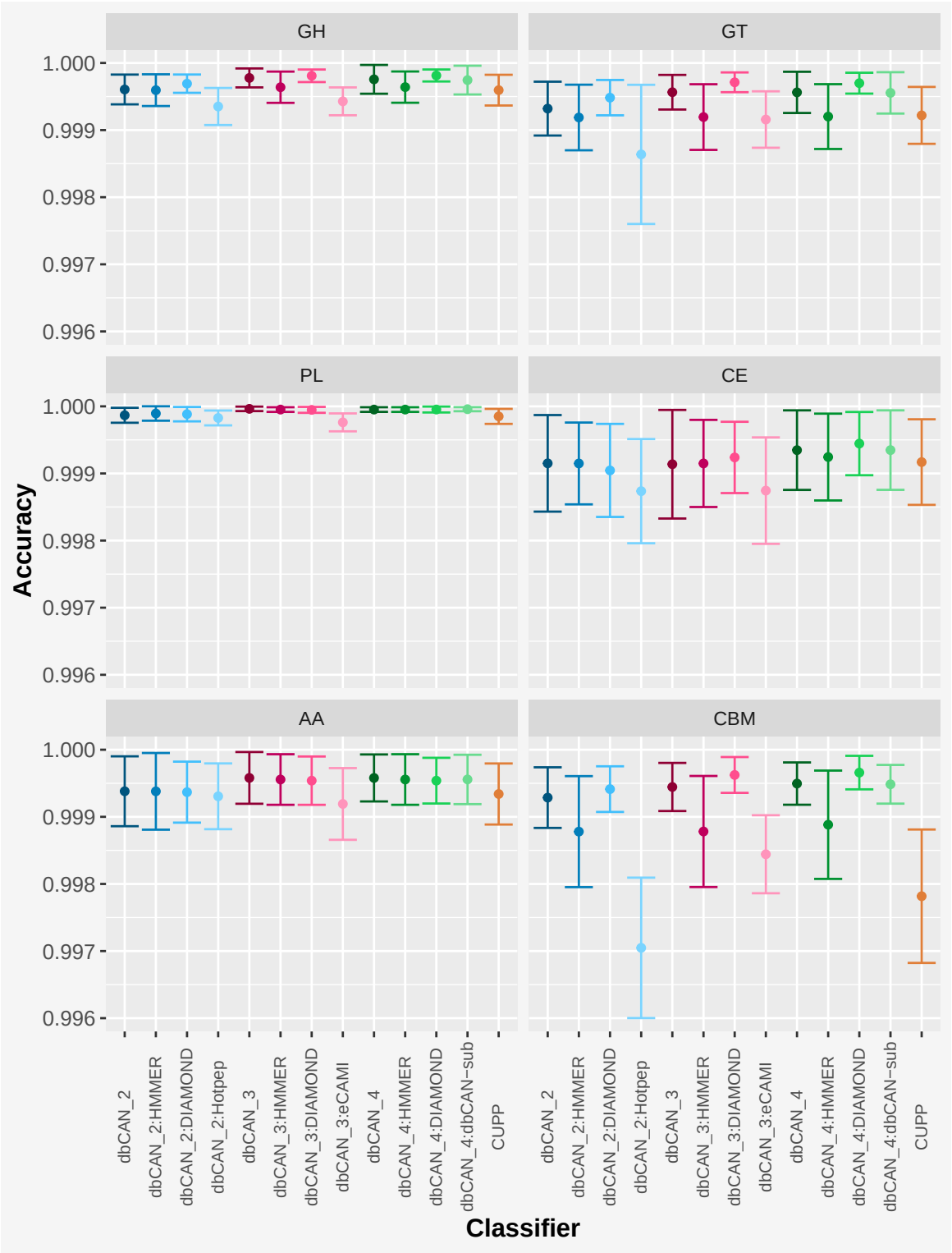



Figure 42: 95% confidence intervals of the accuracy scores calculated for each CAZy family when all test sets are pooled. Where a tool is incorporated into a CAZyme classifier, the tool name is prefixed with the respective CAZyme classifier.

## SI Figure 43: Accuracy of the binary classification of each CAZy family

The accuracy for each CAZy family was calculated, when pooling all test sets.

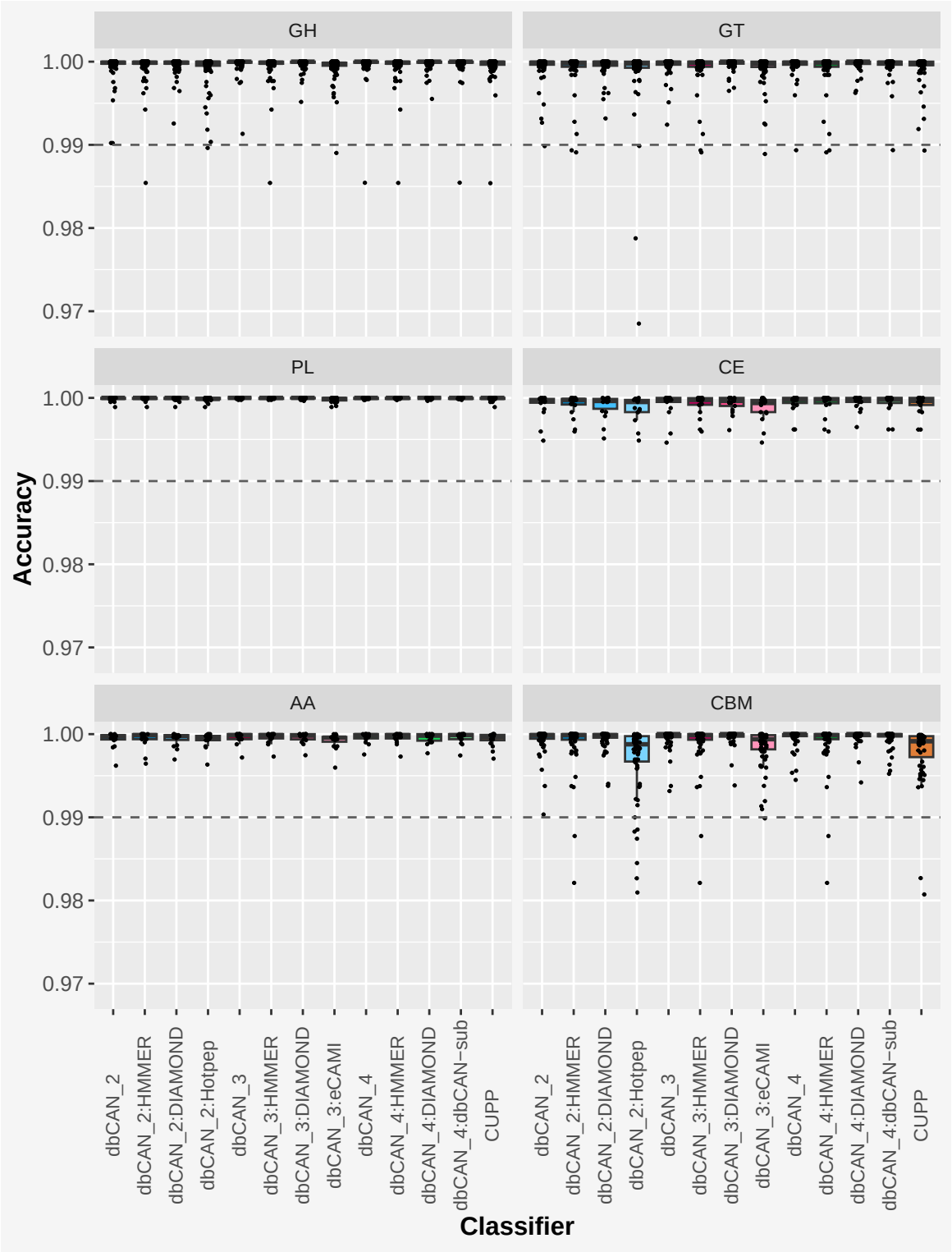



Figure 43: Accuracy of each CAZy family, when pooling all test sets. One-dimensional scatter plot overlaying a box and whisker plot, where each point in the scatter plot represents a CAZy family. Where a tool is incorporated into a CAZyme classifier, the tool name is prefixed with the respective CAZyme classifier.

## 12.6 Sensitivity plotted against specificity

**SI Figure 44: Sensitivity versus specificity of binary classification of each CAZy family from the GH CAZy class**

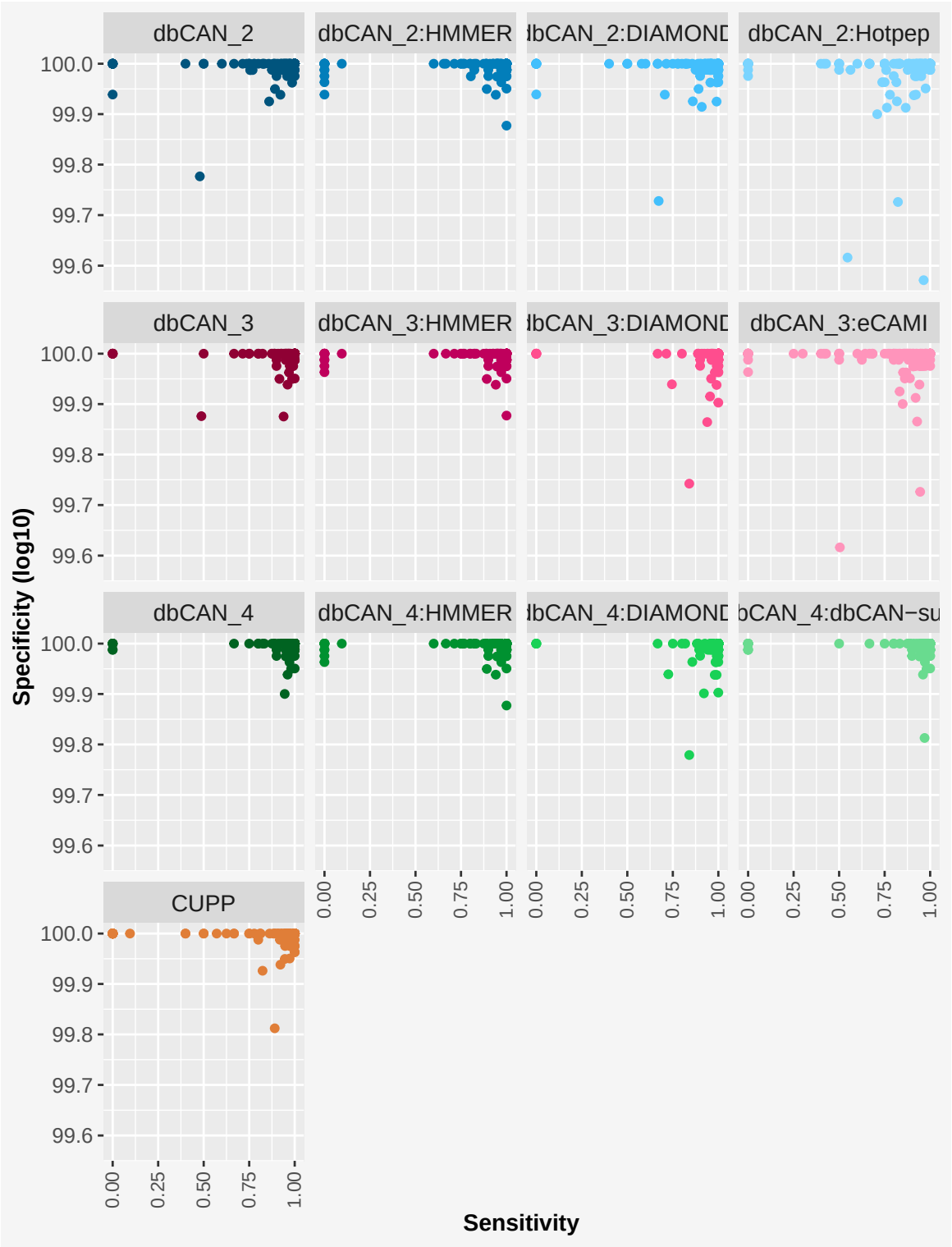



Figure 44: Each family from the GH CAZy class is plotted onto sensitivity versus specificity (the latter plotted on a log10 scale). Where a tool is incorporated into a CAZyme classifier, the tool name is prefixed with the respective CAZyme classifier.

**SI Figure 45: Sensitivity versus specificity of binary classification of each CAZy family from the GT CAZy class**

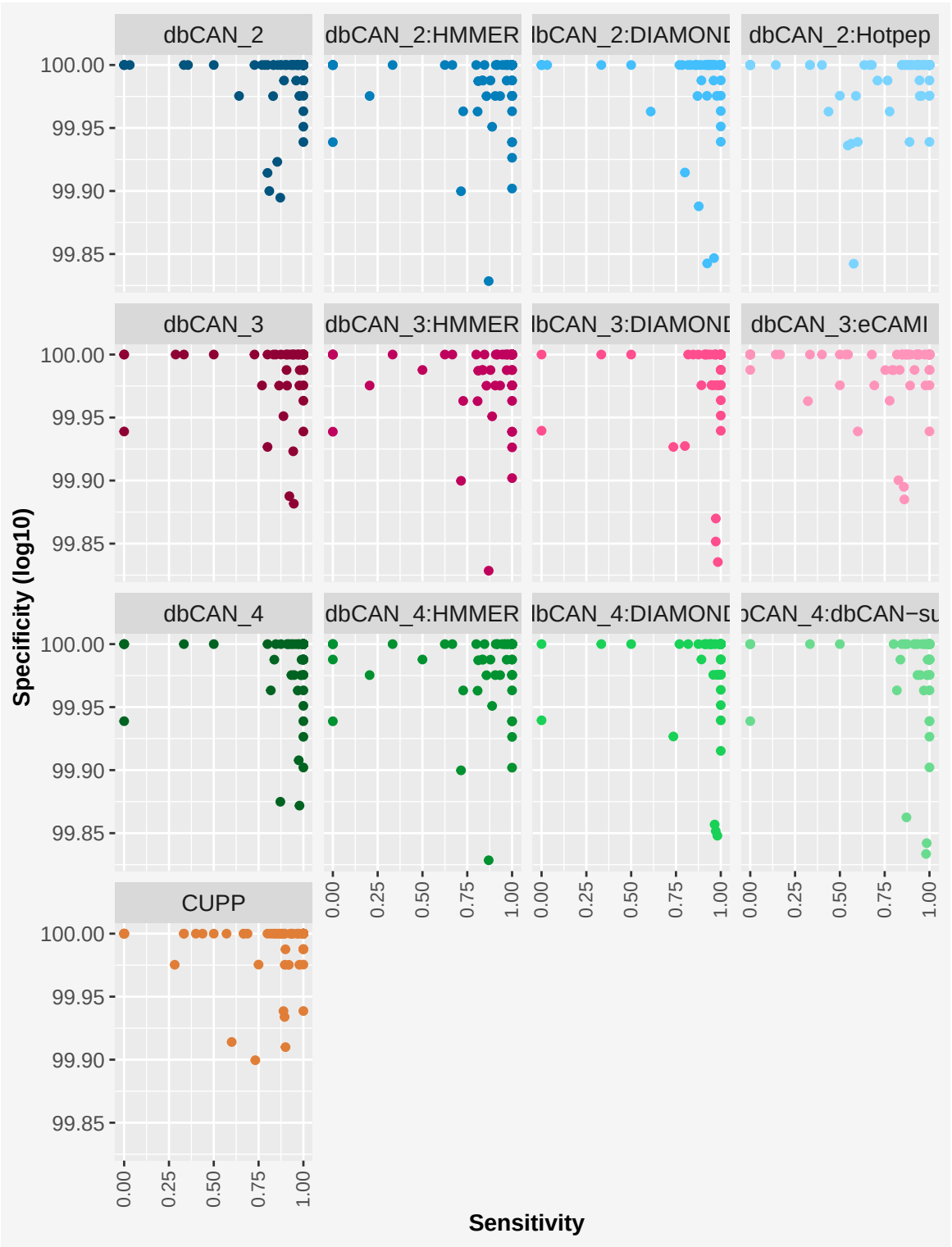



Figure 45: Each family from the GT CAZy class is plotted onto sensitivity versus specificity (the latter plotted on a log10 scale). Where a tool is incorporated into a CAZyme classifier, the tool name is prefixed with the respective CAZyme classifier.

**SI Figure 46: Sensitivity versus specificity of binary classification of each CAZy family from the PL CAZy class**

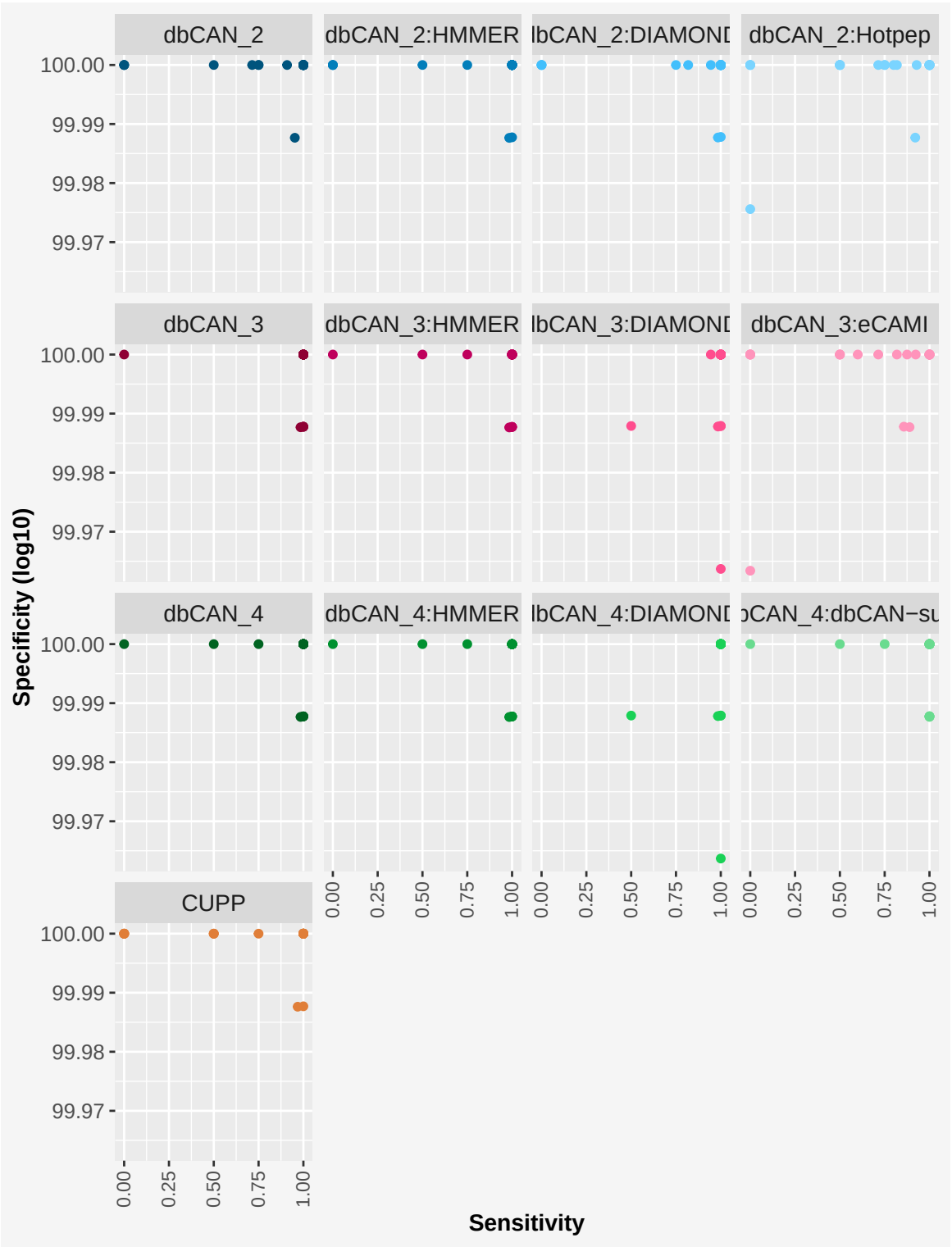



Figure 46: Each family from the PL CAZy class is plotted onto sensitivity versus specificity (the latter plotted on a log10 scale). Where a tool is incorporated into a CAZyme classifier, the tool name is prefixed with the respective CAZyme classifier.

**SI Figure 47: Sensitivity versus specificity of binary classification of each CAZy family from the CE CAZy class**

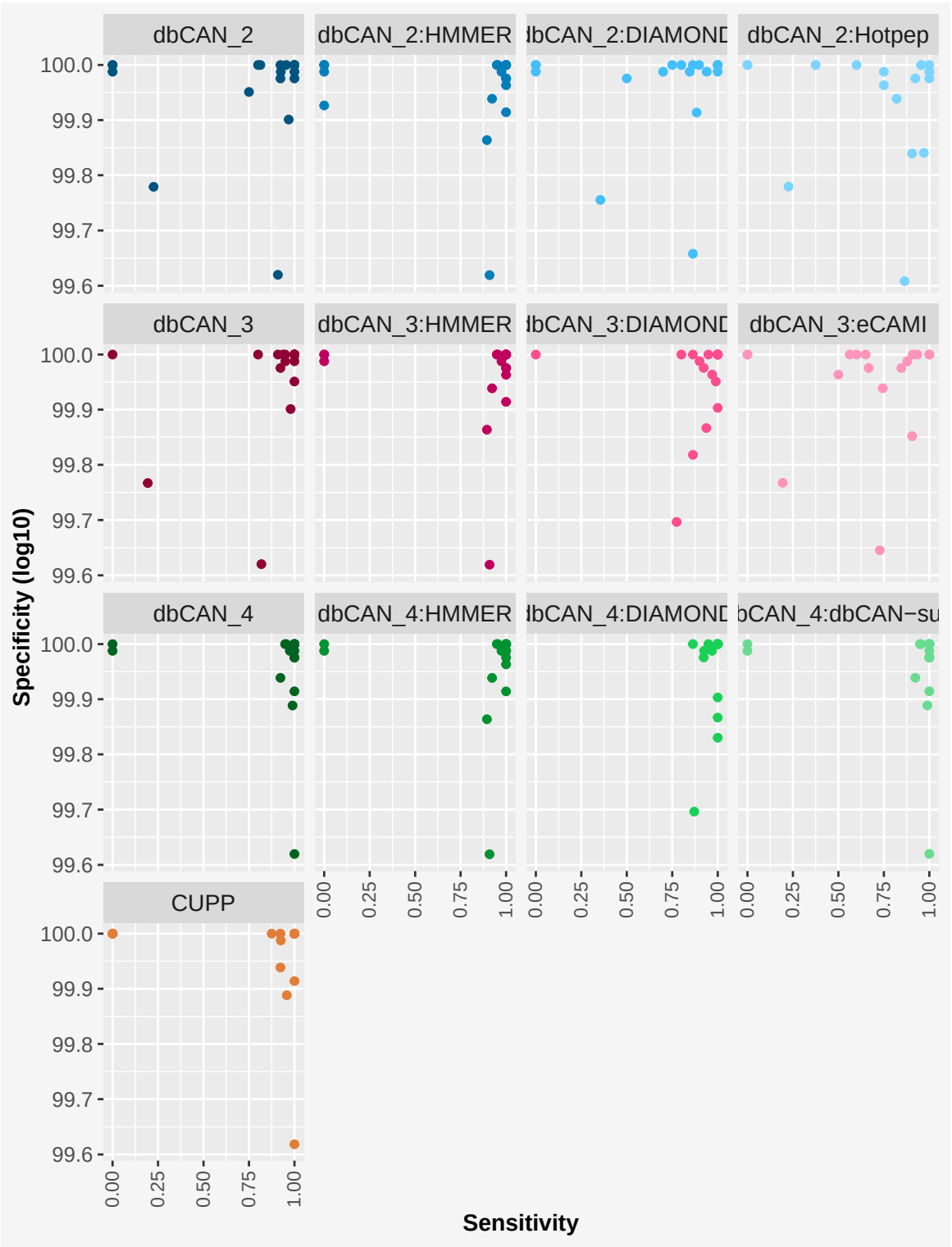



Figure 47: Each family from the CE CAZy class is plotted onto sensitivity versus specificity (the latter plotted on a log10 scale). Where a tool is incorporated into a CAZyme classifier, the tool name is prefixed with the respective CAZyme classifier.

**SI Figure 48: Sensitivity versus specificity of binary classification of each CAZy family from the AA CAZy class**

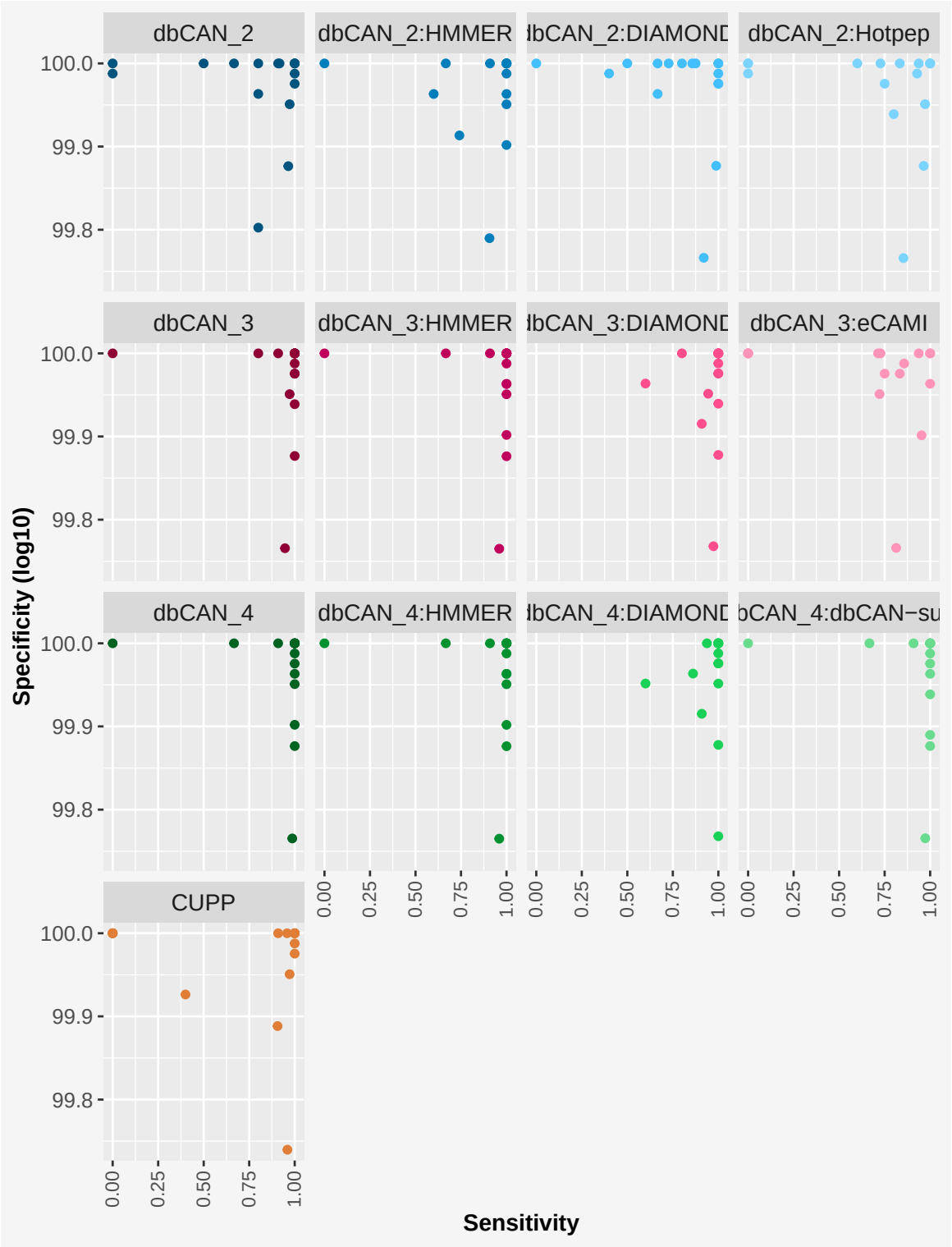



Figure 48: Each family from the AA CAZy class is plotted onto sensitivity versus specificity (the latter plotted on a log10 scale). Where a tool is incorporated into a CAZyme classifier, the tool name is prefixed with the respective CAZyme classifier.

**SI Figure 49: Sensitivity versus specificity of binary classification of each CAZy family from the CBM CAZy class**

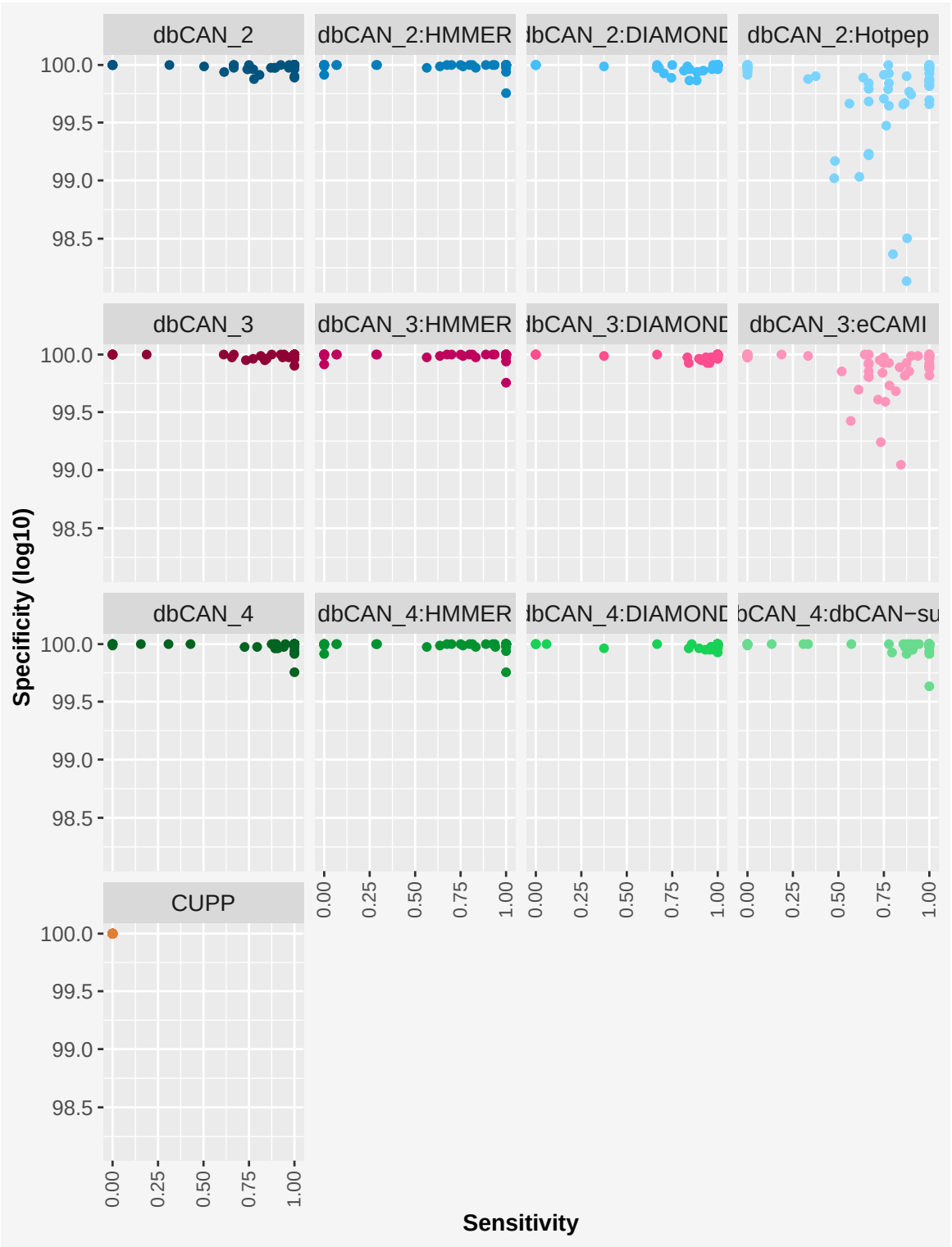



Figure 49: Each family from the CBM CAZy class is plotted onto sensitivity versus specificity (the latter plotted on a log10 scale). Where a tool is incorporated into a CAZyme classifier, the tool name is prefixed with the respective CAZyme classifier.

# 13 Ouput to two-way ANOVAs testing for statistically significant differences in the performances of CAZyme family classification between the CAZyme classes

## SI Table 99: Output of Tukey HSD test for statistically significant differences between the mean f1-score between the CAZyme classes, reporting only tests with a p-value <0.05 (Overleaf).

Reporting only hits where a statistically significant difference (p-value<0.05) is found between the mean F1-scores achieved by a classifier for two different CAZyme classes (i.e. Classifier 1 and Classifier 2 are the same but Class 1 and Class 2 are different). The complete dataset is available in the online repository. Here the pair of CAZy classes, pair of classifiers/tools, the mean difference (Mean Diff), lower and upper 95% confidence interval and the adjusted p-value (Adj.p-value) are reported.

| Class 1 | Class 2 | Classifier 1 | Classifier 2 | MeanDiff | Lower | Upper | Adj.p-value |
|---|---|---|---|---|---|---|---|
| CBM | GH | dbCAN_3:eCAMI | dbCAN_3:eCAMI | -0.29444 | -0.47445 | -0.11444 | 5.70E-08 |
| CBM | GT | dbCAN_3:eCAMI | dbCAN_3:eCAMI | -0.33263 | -0.53454 | -0.13071 | 5.05E-08 |
| CBM | GH | dbCAN_2:Hotpep | dbCAN_2:Hotpep | -0.46701 | -0.64344 | -0.29058 | 3.44E-08 |
| CBM | GT | dbCAN_2:Hotpep | dbCAN_2:Hotpep | -0.45874 | -0.6581 | -0.25937 | 3.44E-08 |
| CBM | PL | dbCAN_2:Hotpep | dbCAN_2:Hotpep | -0.40495 | -0.68818 | -0.12173 | 6.24E-06 |
| CBM | CE | dbCAN_2:Hotpep | dbCAN_2:Hotpep | -0.33855 | -0.66216 | -0.01494 | 0.023636 |
| CBM | AA | dbCAN_2:Hotpep | dbCAN_2:Hotpep | -0.36501 | -0.69664 | -0.03339 | 0.009013 |
| CBM | GH | CUPP | CUPP | -0.80049 | -0.99168 | -0.60931 | 3.44E-08 |
| CBM | GT | CUPP | CUPP | -0.82301 | -1.03518 | -0.61084 | 3.44E-08 |
| CBM | PL | CUPP | CUPP | -0.7307 | -1.02747 | -0.43393 | 3.44E-08 |
| CBM | CE | CUPP | CUPP | -0.68796 | -1.01962 | -0.35631 | 3.44E-08 |
| CBM | AA | CUPP | CUPP | -0.66846 | -1.00794 | -0.32898 | 3.44E-08 |

## SI Table 100: Output of Tukey HSD test for statistically significant differences between the mean accuracy between the CAZyme classes, reporting only tests with a p-value <0.05 (Overleaf).

Reporting only hits where a statistically significant difference (p-value<0.05) is found between the mean accuracy achieved by a classifier for two different CAZyme classes (i.e. Classifier 1 and Classifier 2 are the same but Class 1 and Class 2 are different). The complete dataset is available in the online repository. Here the pair of CAZy classes, pair of classifiers/tools, the mean difference (Mean Diff), lower and upper 95% confidence interval and the adjusted p-value (Adj.p-value) are reported.

| Class 1 | Class 2 | Classifier 1 | Classifier 2 | MeanDiff | Lower | Upper | Adj.p-value |
|---|---|---|---|---|---|---|---|
| CBM | GH | dbCAN_2:Hotpep | dbCAN_2:Hotpep | -0.0023 | -0.00333 | -0.00128 | 3.44E-08 |
| CBM | GT | dbCAN_2:Hotpep | dbCAN_2:Hotpep | -0.00159 | -0.00275 | -0.00043 | 2.36E-05 |
| CBM | PL | dbCAN_2:Hotpep | dbCAN_2:Hotpep | -0.00278 | -0.00442 | -0.00114 | 3.85E-08 |
| CBM | AA | dbCAN_2:Hotpep | dbCAN_2:Hotpep | -0.00226 | -0.00418 | -0.00034 | 0.002144 |
| CBM | GH | CUPP | CUPP | -0.00178 | -0.00289 | -0.00067 | 8.95E-08 |
| CBM | GT | CUPP | CUPP | -0.0014 | -0.00263 | -0.00017 | 0.004261 |
| CBM | PL | CUPP | CUPP | -0.00203 | -0.00375 | -0.00031 | 0.001847 |

## SI Table 101: Output of Tukey HSD test for statistically significant differences between the mean sensitivity between the CAZyme classes, reporting only tests with a p-value <0.05 (Overleaf).

Reporting only hits where a statistically significant difference (p-value<0.05) is found between the mean precision achieved by a classifier for two different CAZyme classes (i.e. Classifier 1 and Classifier 2 are the same but Class 1 and Class 2 are different). The complete dataset is available in the online repository. Here the pair of CAZy classes, pair of classifiers/tools, the mean difference (Mean Diff), lower and upper 95% confidence interval and the adjusted p-value (Adj.p-value) are reported.

| Class 1 | Class 2 | Classifier 1 | Classifier 2 | MeanDiff | Lower | Upper | Adj.p-value |
|---|---|---|---|---|---|---|---|
| CBM | GH | dbCAN_2:Hotpep | dbCAN_2:Hotpep | -0.19413 | -0.3743 | -0.01397 | 0.013711 |
| CBM | GH | CUPP | CUPP | -0.78452 | -0.97975 | -0.58929 | 3.44E-08 |
| CBM | GT | CUPP | CUPP | -0.80394 | -1.0206 | -0.58728 | 3.44E-08 |
| CBM | PL | CUPP | CUPP | -0.70514 | -1.0082 | -0.40209 | 3.44E-08 |
| CBM | CE | CUPP | CUPP | -0.74141 | -1.08009 | -0.40274 | 3.44E-08 |
| CBM | AA | CUPP | CUPP | -0.6941 | -1.04076 | -0.34743 | 3.44E-08 |

## SI Table 102: Output of Tukey HSD test for statistically significant differences between the mean precision between the CAZyme classes, reporting only tests with a p-value <0.05 (Overleaf).

Reporting only hits where a statistically significant difference (p-value<0.05) is found between the mean precision achieved by a classifier for two different CAZyme classes (i.e. Classifier 1 and Classifier 2 are the same but Class 1 and Class 2 are different). The complete dataset is available in the online repository. Here the pair of CAZy classes, pair of classifiers/tools, the mean difference (Mean Diff), lower and upper 95% confidence interval and the adjusted p-value (Adj.p-value) are reported.

| Class 1 | Class 2 | Classifier 1 | Classifier 2 | MeanDiff | Lower | Upper | Adj.p-value |
|---|---|---|---|---|---|---|---|
| CBM | GH | dbCAN_3:eCAMI | dbCAN_3:eCAMI | -0.39005 | -0.57336 | -0.20674 | 3.44E-08 |
| CBM | GT | dbCAN_3:eCAMI | dbCAN_3:eCAMI | -0.43828 | -0.6439 | -0.23265 | 3.44E-08 |
| CBM | PL | dbCAN_3:eCAMI | dbCAN_3:eCAMI | -0.35 | -0.64529 | -0.05471 | 0.001736 |
| CBM | GH | dbCAN_2:Hotpep | dbCAN_2:Hotpep | -0.57613 | -0.7558 | -0.39646 | 3.44E-08 |
| CBM | GT | dbCAN_2:Hotpep | dbCAN_2:Hotpep | -0.58387 | -0.7869 | -0.38085 | 3.44E-08 |
| CBM | PL | dbCAN_2:Hotpep | dbCAN_2:Hotpep | -0.55382 | -0.84225 | -0.26539 | 3.44E-08 |
| CBM | CE | dbCAN_2:Hotpep | dbCAN_2:Hotpep | -0.41857 | -0.74813 | -0.08902 | 0.000282 |
| CBM | AA | dbCAN_2:Hotpep | dbCAN_2:Hotpep | -0.442 | -0.77971 | -0.10428 | 0.000117 |
| CBM | GH | CUPP | CUPP | -0.83306 | -1.02775 | -0.63836 | 3.44E-08 |
| CBM | GT | CUPP | CUPP | -0.87849 | -1.09456 | -0.66243 | 3.44E-08 |
| CBM | PL | CUPP | CUPP | -0.77962 | -1.08184 | -0.47739 | 3.44E-08 |
| CBM | CE | CUPP | CUPP | -0.66872 | -1.00647 | -0.33098 | 3.44E-08 |
| CBM | AA | CUPP | CUPP | -0.65106 | -0.99677 | -0.30535 | 3.44E-08 |

# 14 Ouput to two-way ANOVAs testing for statistically significant differences in the performances of CAZyme family classification between the classifiers

## SI table 103: Output of Tukey HSD test for statistically significant differences between the mean F1-score between CAZyme classifiers classifying CAZyme families from the same CAZyme class, reporting only tests with a p-value <0.05 (Overleaf)

Reporting only hits where a statistically significant difference (p-value<0.05) is found between the mean F1-scores achieved by two different classifiers for same class (i.e. Classifier 1 and Classifier 2 are different but Class 1 and Class 2 are the same). The complete dataset is available in the online repository. Here the pair of CAZy classes, pair of classifiers/tools, the mean difference (Mean Diff), lower and upper 95% confidence interval and the adjusted p-value (Adj.p-value) are reported.

| Class 1 | Class 2 | Classifier 1 | Classifier 2 | MeanDiff | Lower | Upper | Adj.p-value |
|---|---|---|---|---|---|---|---|
| GH | GH | dbCAN_3:eCAMI | dbCAN_3 | -0.17799 | -0.32293 | -0.03305 | 0.000709 |
| GH | GH | dbCAN_3:eCAMI | dbCAN_3:DIAMOND | -0.18244 | -0.32737 | -0.0375 | 0.000363 |
| GH | GH | dbCAN_4 | dbCAN_3:eCAMI | 0.159761 | 0.015357 | 0.304164 | 0.008116 |
| GH | GH | dbCAN_4:DIAMOND | dbCAN_3:eCAMI | 0.188177 | 0.04324 | 0.333114 | 0.000149 |
| GH | GH | dbCAN_4:sub | dbCAN_3:eCAMI | 0.147597 | 0.003455 | 0.291738 | 0.034157 |
| GH | GH | CUPP | dbCAN_4:DIAMOND | -0.14797 | -0.29344 | -0.0025 | 0.03813 |
| CBM | CBM | dbCAN_2:Hotpep | dbCAN_2 | -0.44955 | -0.66558 | -0.23351 | 3.44E-08 |
| CBM | CBM | dbCAN_3:eCAMI | dbCAN_2 | -0.32529 | -0.54426 | -0.10633 | 1.50E-06 |
| CBM | CBM | CUPP | dbCAN_2 | -0.79113 | -1.01904 | -0.56323 | 3.44E-08 |
| CBM | CBM | dbCAN_2:Hotpep | dbCAN_2:HMMER | -0.35844 | -0.5734 | -0.14348 | 4.33E-08 |
| CBM | CBM | dbCAN_3:eCAMI | dbCAN_2:HMMER | -0.23419 | -0.45209 | -0.01628 | 0.014406 |
| CBM | CBM | CUPP | dbCAN_2:HMMER | -0.70003 | -0.92692 | -0.47314 | 3.44E-08 |
| CBM | CBM | dbCAN_2:Hotpep | dbCAN_2:DIAMOND | -0.45563 | -0.67167 | -0.2396 | 3.44E-08 |
| CBM | CBM | dbCAN_3:eCAMI | dbCAN_2:DIAMOND | -0.33138 | -0.55034 | -0.11241 | 7.31E-07 |
| CBM | CBM | CUPP | dbCAN_2:DIAMOND | -0.79722 | -1.02513 | -0.56931 | 3.44E-08 |
| CBM | CBM | dbCAN_3 | dbCAN_2:Hotpep | 0.49177 | 0.275734 | 0.707807 | 3.44E-08 |
| CBM | CBM | dbCAN_3:HMMER | dbCAN_2:Hotpep | 0.3763 | 0.16134 | 0.591261 | 3.51E-08 |
| CBM | CBM | dbCAN_3:DIAMOND | dbCAN_2:Hotpep | 0.513495 | 0.297458 | 0.729532 | 3.44E-08 |
| CBM | CBM | dbCAN_4 | dbCAN_2:Hotpep | 0.472783 | 0.257822 | 0.687744 | 3.44E-08 |
| CBM | CBM | dbCAN_4:HMMER | dbCAN_2:Hotpep | 0.393274 | 0.178313 | 0.608234 | 3.44E-08 |
| CBM | CBM | dbCAN_4:DIAMOND | dbCAN_2:Hotpep | 0.516535 | 0.300498 | 0.732572 | 3.44E-08 |
| CBM | CBM | dbCAN_4:sub | dbCAN_2:Hotpep | 0.441614 | 0.226653 | 0.656575 | 3.44E-08 |
| CBM | CBM | CUPP | dbCAN_2:Hotpep | -0.34159 | -0.55762 | -0.12555 | 1.42E-07 |
| CBM | CBM | dbCAN_3:eCAMI | dbCAN_3 | -0.36752 | -0.58648 | -0.14855 | 4.08E-08 |
| CBM | CBM | CUPP | dbCAN_3 | -0.83336 | -1.06126 | -0.60545 | 3.44E-08 |
| CBM | CBM | dbCAN_3:eCAMI | dbCAN_3:HMMER | -0.25205 | -0.46995 | -0.03414 | 0.003089 |
| CBM | CBM | CUPP | dbCAN_3:HMMER | -0.71789 | -0.94477 | -0.491 | 3.44E-08 |
| CBM | CBM | dbCAN_3:eCAMI | dbCAN_3:DIAMOND | -0.38924 | -0.60821 | -0.17027 | 3.47E-08 |
| CBM | CBM | CUPP | dbCAN_3:DIAMOND | -0.85508 | -1.08299 | -0.62717 | 3.44E-08 |
| CBM | CBM | dbCAN_4 | dbCAN_3:eCAMI | 0.348528 | 0.130624 | 0.566432 | 9.87E-08 |
| CBM | CBM | dbCAN_4:HMMER | dbCAN_3:eCAMI | 0.269019 | 0.051115 | 0.486923 | 0.000616 |
| CBM | CBM | dbCAN_4:DIAMOND | dbCAN_3:eCAMI | 0.39228 | 0.173315 | 0.611246 | 3.46E-08 |
| CBM | CBM | dbCAN_4:sub | dbCAN_3:eCAMI | 0.317359 | 0.099455 | 0.535263 | 3.17E-06 |
| CBM | CBM | CUPP | dbCAN_3:eCAMI | -0.46584 | -0.68481 | -0.24688 | 3.44E-08 |
| CBM | CBM | CUPP | dbCAN_4 | -0.81437 | -1.04126 | -0.58748 | 3.44E-08 |
| CBM | CBM | CUPP | dbCAN_4:HMMER | -0.73486 | -0.96175 | -0.50797 | 3.44E-08 |
| CBM | CBM | CUPP | dbCAN_4:DIAMOND | -0.85812 | -1.08603 | -0.63021 | 3.44E-08 |
| CBM | CBM | CUPP | dbCAN_4:sub | -0.7832 | -1.01009 | -0.55631 | 3.44E-08 |

## SI table 104: Output of Tukey HSD test for statistically significant differences between the mean sensitivity between CAZyme classifiers classifying CAZyme families from the same CAZyme class, reporting only tests with a p-value <0.05 (Overleaf)

Reporting only hits where a statistically significant difference (p-value<0.05) is found between the mean sensitivity achieved by two different classifiers for same class (i.e. Classifier 1 and Classifier 2 are different but Class 1 and Class 2 are the same). The complete dataset is available in the online repository. Here the pair of CAZy classes, pair of classifiers/tools, the mean difference (Mean Diff), lower and upper 95% confidence interval and the adjusted p-value (Adj.p-value) are reported.

| Class 1 | Class 2 | Classifier 1 | Classifier 2 | MeanDiff | Lower | Upper | Adj.p-value |
|---|---|---|---|---|---|---|---|
| GH | GH | dbCAN_3:DIAMOND | dbCAN_2:Hotpep | 0.161969 | 0.013963 | 0.309974 | 0.010124 |
| GH | GH | dbCAN_4:DIAMOND | dbCAN_2:Hotpep | 0.166907 | 0.018901 | 0.314912 | 0.005428 |
| GH | GH | dbCAN_3:eCAMI | dbCAN_3 | -0.20248 | -0.35049 | -0.05447 | 2.88E-05 |
| GH | GH | dbCAN_3:eCAMI | dbCAN_3:HMMER | -0.15389 | -0.30082 | -0.00696 | 0.023155 |
| GH | GH | dbCAN_3:eCAMI | dbCAN_3:DIAMOND | -0.21828 | -0.36629 | -0.07028 | 1.97E-06 |
| GH | GH | CUPP | dbCAN_3:DIAMOND | -0.15944 | -0.30799 | -0.01089 | 0.014768 |
| GH | GH | dbCAN_4 | dbCAN_3:eCAMI | 0.194044 | 0.046584 | 0.341505 | 9.94E-05 |
| GH | GH | dbCAN_4:HMMER | dbCAN_3:eCAMI | 0.161083 | 0.014154 | 0.308013 | 0.009762 |
| GH | GH | dbCAN_4:DIAMOND | dbCAN_3:eCAMI | 0.223222 | 0.075217 | 0.371228 | 8.35E-07 |
| GH | GH | dbCAN_4:sub | dbCAN_3:eCAMI | 0.183917 | 0.036724 | 0.33111 | 0.000445 |
| GH | GH | CUPP | dbCAN_4:DIAMOND | -0.16438 | -0.31293 | -0.01583 | 0.008083 |
| CBM | CBM | dbCAN_2:Hotpep | dbCAN_2 | -0.22199 | -0.4426 | -0.00138 | 0.045345 |
| CBM | CBM | CUPP | dbCAN_2 | -0.80985 | -1.04258 | -0.57712 | 3.44E-08 |
| CBM | CBM | CUPP | dbCAN_2:HMMER | -0.69649 | -0.92818 | -0.4648 | 3.44E-08 |
| CBM | CBM | dbCAN_2:Hotpep | dbCAN_2:DIAMOND | -0.23242 | -0.45303 | -0.01181 | 0.020841 |
| CBM | CBM | CUPP | dbCAN_2:DIAMOND | -0.82029 | -1.05302 | -0.58756 | 3.44E-08 |
| CBM | CBM | dbCAN_3 | dbCAN_2:Hotpep | 0.258789 | 0.038178 | 0.479399 | 0.002226 |
| CBM | CBM | dbCAN_3:DIAMOND | dbCAN_2:Hotpep | 0.289315 | 0.068705 | 0.509926 | 0.00011 |
| CBM | CBM | dbCAN_4 | dbCAN_2:Hotpep | 0.245342 | 0.02583 | 0.464853 | 0.006567 |
| CBM | CBM | dbCAN_4:DIAMOND | dbCAN_2:Hotpep | 0.29327 | 0.07266 | 0.513881 | 7.26E-05 |
| CBM | CBM | CUPP | dbCAN_2:Hotpep | -0.58786 | -0.80847 | -0.36725 | 3.44E-08 |
| CBM | CBM | dbCAN_3:eCAMI | dbCAN_3 | -0.24181 | -0.46541 | -0.01821 | 0.012783 |
| CBM | CBM | CUPP | dbCAN_3 | -0.84665 | -1.07938 | -0.61392 | 3.44E-08 |
| CBM | CBM | CUPP | dbCAN_3:HMMER | -0.71434 | -0.94603 | -0.48265 | 3.44E-08 |
| CBM | CBM | dbCAN_3:eCAMI | dbCAN_3:DIAMOND | -0.27234 | -0.49594 | -0.04874 | 0.000878 |
| CBM | CBM | CUPP | dbCAN_3:DIAMOND | -0.87718 | -1.10991 | -0.64445 | 3.44E-08 |
| CBM | CBM | dbCAN_4 | dbCAN_3:eCAMI | 0.228364 | 0.005847 | 0.450881 | 0.032897 |
| CBM | CBM | dbCAN_4:DIAMOND | dbCAN_3:eCAMI | 0.276293 | 0.052692 | 0.499894 | 0.000602 |
| CBM | CBM | CUPP | dbCAN_3:eCAMI | -0.60484 | -0.82844 | -0.38124 | 3.44E-08 |
| CBM | CBM | CUPP | dbCAN_4 | -0.8332 | -1.06489 | -0.60151 | 3.44E-08 |
| CBM | CBM | CUPP | dbCAN_4:HMMER | -0.73115 | -0.96284 | -0.49946 | 3.44E-08 |
| CBM | CBM | CUPP | dbCAN_4:DIAMOND | -0.88113 | -1.11386 | -0.6484 | 3.44E-08 |
| CBM | CBM | CUPP | dbCAN_4:sub | -0.80117 | -1.03286 | -0.56948 | 3.44E-08 |

## SI table 105: Output of Tukey HSD test for statistically significant differences between the mean accuracy between CAZyme classifiers classifying CAZyme families from the same CAZyme class, reporting only tests with a p-value <0.05 (Overleaf)

Reporting only hits where a statistically significant difference (p-value<0.05) is found between the mean accuracy achieved by two different classifiers for same class (i.e. Classifier 1 and Classifier 2 are different but Class 1 and Class 2 are the same). The complete dataset is available in the online repository. Here the pair of CAZy classes, pair of classifiers/tools, the mean difference (Mean Diff), lower and upper 95% confidence interval and the adjusted p-value (Adj.p-value) are reported.

| Class 1 | Class 2 | Classifier 1 | Classifier 2 | MeanDiff | Lower | Upper | Adj.p-value |
|---|---|---|---|---|---|---|---|
| CBM | CBM | dbCAN_2:Hotpep | dbCAN_2 | -0.00224 | -0.00349 | -0.00099 | 3.46E-08 |
| CBM | CBM | CUPP | dbCAN_2 | -0.00147 | -0.00279 | -0.00015 | 0.00727 |
| CBM | CBM | dbCAN_2:Hotpep | dbCAN_2:HMMER | -0.00173 | -0.00298 | -0.00049 | 1.68E-05 |
| CBM | CBM | dbCAN_2:Hotpep | dbCAN_2:DIAMOND | -0.00237 | -0.00362 | -0.00111 | 3.44E-08 |
| CBM | CBM | CUPP | dbCAN_2:DIAMOND | -0.0016 | -0.00292 | -0.00027 | 0.001083 |
| CBM | CBM | dbCAN_3 | dbCAN_2:Hotpep | 0.002397 | 0.001145 | 0.003649 | 3.44E-08 |
| CBM | CBM | dbCAN_3:HMMER | dbCAN_2:Hotpep | 0.001734 | 0.000488 | 0.00298 | 1.62E-05 |
| CBM | CBM | dbCAN_3:DIAMOND | dbCAN_2:Hotpep | 0.002576 | 0.001324 | 0.003828 | 3.44E-08 |
| CBM | CBM | dbCAN_3:eCAMI | dbCAN_2:Hotpep | 0.001395 | 0.000197 | 0.002592 | 0.002622 |
| CBM | CBM | dbCAN_4 | dbCAN_2:Hotpep | 0.002448 | 0.001202 | 0.003693 | 3.44E-08 |
| CBM | CBM | dbCAN_4:HMMER | dbCAN_2:Hotpep | 0.001834 | 0.000589 | 0.00308 | 2.10E-06 |
| CBM | CBM | dbCAN_4:DIAMOND | dbCAN_2:Hotpep | 0.002611 | 0.001359 | 0.003863 | 3.44E-08 |
| CBM | CBM | dbCAN_4:sub | dbCAN_2:Hotpep | 0.002437 | 0.001191 | 0.003683 | 3.44E-08 |
| CBM | CBM | CUPP | dbCAN_3 | -0.00163 | -0.00295 | -0.00031 | 0.000648 |
| CBM | CBM | CUPP | dbCAN_3:DIAMOND | -0.00181 | -0.00313 | -0.00049 | 2.89E-05 |
| CBM | CBM | CUPP | dbCAN_4 | -0.00168 | -0.00299 | -0.00036 | 0.000246 |
| CBM | CBM | CUPP | dbCAN_4:DIAMOND | -0.00184 | -0.00316 | -0.00052 | 1.51E-05 |
| CBM | CBM | CUPP | dbCAN_4:sub | -0.00167 | -0.00298 | -0.00035 | 0.000294 |

## SI table 106: Output of Tukey HSD test for statistically significant differences between the mean precision between CAZyme classifiers classifying CAZyme families from the same CAZyme class, reporting only tests with a p-value <0.05 (Overleaf)

Reporting only hits where a statistically significant difference (p-value<0.05) is found between the mean precision achieved by two different classifiers for the same class (i.e. Classifier 1 and Classifier 2 are different but Class 1 and Class 2 are the same). The complete dataset is available in the online repository. Here the pair of CAZy classes, pair of classifiers/tools, the mean difference (Mean Diff), lower and upper 95% confidence interval and the adjusted p-value (Adj.p-value) are reported.

| Class 1 | Class 2 | Classifier 1 | Classifier 2 | MeanDiff | Lower | Upper | Adj.p-value |
|---|---|---|---|---|---|---|---|
| CBM | CBM | dbCAN_2:Hotpep | dbCAN_2 | -0.53376 | -0.75377 | -0.31376 | 3.44E-08 |
| CBM | CBM | dbCAN_3:eCAMI | dbCAN_2 | -0.38407 | -0.60705 | -0.16108 | 3.61E-08 |
| CBM | CBM | CUPP | dbCAN_2 | -0.81257 | -1.04466 | -0.58047 | 3.44E-08 |
| CBM | CBM | dbCAN_2:Hotpep | dbCAN_2:HMMER | -0.48902 | -0.70793 | -0.27011 | 3.44E-08 |
| CBM | CBM | dbCAN_3:eCAMI | dbCAN_2:HMMER | -0.33933 | -0.56123 | -0.11742 | 4.89E-07 |
| CBM | CBM | CUPP | dbCAN_2:HMMER | -0.76783 | -0.99888 | -0.53677 | 3.44E-08 |
| CBM | CBM | dbCAN_2:Hotpep | dbCAN_2:DIAMOND | -0.51914 | -0.73914 | -0.29913 | 3.44E-08 |
| CBM | CBM | dbCAN_3:eCAMI | dbCAN_2:DIAMOND | -0.36944 | -0.59243 | -0.14645 | 4.66E-08 |
| CBM | CBM | CUPP | dbCAN_2:DIAMOND | -0.79794 | -1.03003 | -0.56585 | 3.44E-08 |
| CBM | CBM | dbCAN_3 | dbCAN_2:Hotpep | 0.576379 | 0.356373 | 0.796384 | 3.44E-08 |
| CBM | CBM | dbCAN_3:HMMER | dbCAN_2:Hotpep | 0.50688 | 0.287971 | 0.72579 | 3.44E-08 |
| CBM | CBM | dbCAN_3:DIAMOND | dbCAN_2:Hotpep | 0.570222 | 0.350216 | 0.790227 | 3.44E-08 |
| CBM | CBM | dbCAN_4 | dbCAN_2:Hotpep | 0.566262 | 0.347352 | 0.785171 | 3.44E-08 |
| CBM | CBM | dbCAN_4:HMMER | dbCAN_2:Hotpep | 0.524023 | 0.305114 | 0.742932 | 3.44E-08 |
| CBM | CBM | dbCAN_4:DIAMOND | dbCAN_2:Hotpep | 0.588807 | 0.368801 | 0.808812 | 3.44E-08 |
| CBM | CBM | dbCAN_4:sub | dbCAN_2:Hotpep | 0.542443 | 0.323533 | 0.761352 | 3.44E-08 |
| CBM | CBM | CUPP | dbCAN_2:Hotpep | -0.2788 | -0.49881 | -0.0588 | 0.000301 |
| CBM | CBM | dbCAN_3:eCAMI | dbCAN_3 | -0.42668 | -0.64967 | -0.20369 | 3.44E-08 |
| CBM | CBM | CUPP | dbCAN_3 | -0.85518 | -1.08728 | -0.62309 | 3.44E-08 |
| CBM | CBM | dbCAN_3:eCAMI | dbCAN_3:HMMER | -0.35718 | -0.57909 | -0.13528 | 8.25E-08 |
| CBM | CBM | CUPP | dbCAN_3:HMMER | -0.78568 | -1.01674 | -0.55463 | 3.44E-08 |
| CBM | CBM | dbCAN_3:eCAMI | dbCAN_3:DIAMOND | -0.42052 | -0.64351 | -0.19754 | 3.44E-08 |
| CBM | CBM | CUPP | dbCAN_3:DIAMOND | -0.84903 | -1.08112 | -0.61693 | 3.44E-08 |
| CBM | CBM | dbCAN_4 | dbCAN_3:eCAMI | 0.416565 | 0.194658 | 0.638471 | 3.44E-08 |
| CBM | CBM | dbCAN_4:HMMER | dbCAN_3:eCAMI | 0.374326 | 0.15242 | 0.596232 | 3.94E-08 |
| CBM | CBM | dbCAN_4:DIAMOND | dbCAN_3:eCAMI | 0.439109 | 0.216122 | 0.662097 | 3.44E-08 |
| CBM | CBM | dbCAN_4:sub | dbCAN_3:eCAMI | 0.392746 | 0.170839 | 0.614652 | 3.48E-08 |
| CBM | CBM | CUPP | dbCAN_3:eCAMI | -0.4285 | -0.65149 | -0.20551 | 3.44E-08 |
| CBM | CBM | CUPP | dbCAN_4 | -0.84507 | -1.07612 | -0.61401 | 3.44E-08 |
| CBM | CBM | CUPP | dbCAN_4:HMMER | -0.80283 | -1.03388 | -0.57177 | 3.44E-08 |
| CBM | CBM | CUPP | dbCAN_4:DIAMOND | -0.86761 | -1.0997 | -0.63552 | 3.44E-08 |
| CBM | CBM | CUPP | dbCAN_4:sub | -0.82125 | -1.0523 | -0.59019 | 3.44E-08 |

## SI table 107: The standard deviations around the mean F1-score of CAZyme family classification

Table 1: Range of the standard deviation around the mean F1-score of CAZyme family classification

| CAZyme classifier | GH | GT | PL | CE | AA | CBM |
|---|---|---|---|---|---|---|
| dbCAN_2 | 0.262 | 0.261 | 0.339 | 0.377 | 0.320 | 0.326 |
| dbCAN_2:HMMER | 0.310 | 0.298 | 0.344 | 0.417 | 0.259 | 0.407 |
| dbCAN_2:DIAMOND | 0.262 | 0.259 | 0.340 | 0.365 | 0.248 | 0.312 |
| dbCAN_2:Hotpep | 0.315 | 0.307 | 0.361 | 0.322 | 0.371 | 0.316 |
| dbCAN_3 | 0.193 | 0.223 | 0.216 | 0.299 | 0.244 | 0.306 |
| dbCAN_3:HMMER | 0.275 | 0.250 | 0.222 | 0.367 | 0.249 | 0.397 |
| dbCAN_3:DIAMOND | 0.191 | 0.190 | 0.131 | 0.264 | 0.128 | 0.293 |
| dbCAN_3:eCAMI | 0.353 | 0.310 | 0.396 | 0.332 | 0.395 | 0.345 |
| dbCAN_4 | 0.235 | 0.239 | 0.222 | 0.323 | 0.248 | 0.341 |
| dbCAN_4:HMMER | 0.264 | 0.267 | 0.222 | 0.322 | 0.249 | 0.386 |
| dbCAN_4:DIAMOND | 0.173 | 0.188 | 0.131 | 0.150 | 0.137 | 0.292 |
| dbCAN_4:dbCAN-sub | 0.248 | 0.240 | 0.222 | 0.323 | 0.249 | 0.359 |
| CUPP | 0.360 | 0.288 | 0.409 | 0.417 | 0.432 | 0.000 |
| Mean | 0.265 | 0.256 | 0.274 | 0.329 | 0.272 | 0.314 |
| Standard Deviation | 0.057 | 0.038 | 0.092 | 0.067 | 0.086 | 0.098 |

# 15 Multilabel classification of CAZy families

## SI Table 108: Multilabel classification of CAZy families

| Rand Index | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Bacteria | | | | All | | | | Eukaryote | | | |
| Classifier | Lower CI | Mean | SD | Upper CI | Lower CI | Mean | SD | Upper CI | Lower CI | Mean | SD | Upper CI |
| dbCAN_2 | 0.977 | 0.978 | 0.082 | 0.980 | 0.976 | 0.977 | 0.086 | 0.978 | 0.973 | 0.975 | 0.088 | 0.977 |
| dbCAN_2:HMMER | 0.968 | 0.970 | 0.096 | 0.972 | 0.968 | 0.969 | 0.097 | 0.971 | 0.967 | 0.969 | 0.098 | 0.971 |
| dbCAN_2:DIAMOND | 0.978 | 0.980 | 0.080 | 0.982 | 0.978 | 0.979 | 0.082 | 0.980 | 0.976 | 0.978 | 0.084 | 0.980 |
| dbCAN_2:Hotpep | 0.944 | 0.947 | 0.125 | 0.949 | 0.944 | 0.946 | 0.125 | 0.948 | 0.944 | 0.946 | 0.125 | 0.949 |
| dbCAN_3 | 0.982 | 0.984 | 0.072 | 0.986 | 0.983 | 0.984 | 0.072 | 0.985 | 0.982 | 0.984 | 0.072 | 0.985 |
| dbCAN_3:HMMER | 0.969 | 0.971 | 0.094 | 0.973 | 0.970 | 0.971 | 0.094 | 0.973 | 0.970 | 0.972 | 0.093 | 0.974 |
| dbCAN_3:DIAMOND | 0.985 | 0.987 | 0.066 | 0.988 | 0.986 | 0.987 | 0.064 | 0.988 | 0.987 | 0.988 | 0.062 | 0.990 |
| dbCAN_3:eCAMI | 0.962 | 0.964 | 0.105 | 0.967 | 0.959 | 0.961 | 0.109 | 0.963 | 0.955 | 0.958 | 0.113 | 0.960 |
| dbCAN_4 | 0.986 | 0.988 | 0.063 | 0.989 | 0.985 | 0.986 | 0.068 | 0.987 | 0.982 | 0.983 | 0.073 | 0.985 |
| dbCAN_4:HMMER | 0.971 | 0.973 | 0.091 | 0.975 | 0.971 | 0.973 | 0.092 | 0.974 | 0.970 | 0.972 | 0.093 | 0.974 |
| dbCAN_4:DIAMOND | 0.985 | 0.986 | 0.067 | 0.988 | 0.987 | 0.988 | 0.063 | 0.989 | 0.989 | 0.990 | 0.058 | 0.991 |
| dbCAN_4:dbCAN-sub | 0.987 | 0.988 | 0.061 | 0.990 | 0.984 | 0.985 | 0.069 | 0.986 | 0.980 | 0.982 | 0.076 | 0.983 |
| CUPP | 0.957 | 0.959 | 0.110 | 0.962 | 0.958 | 0.960 | 0.109 | 0.961 | 0.958 | 0.960 | 0.109 | 0.963 |
| **Adjusted Rand Index** | | | | | | | | | | | | |
| | Bacteria | | | | All | | | | Eukaryote | | | |
| Classifier | Lower CI | Mean | SD | Upper CI | Lower CI | Mean | SD | Upper CI | Lower CI | Mean | SD | Upper CI |
| dbCAN_2 | 0.933 | 0.938 | 0.237 | 0.944 | 0.930 | 0.934 | 0.245 | 0.938 | 0.924 | 0.929 | 0.254 | 0.935 |
| dbCAN_2:HMMER | 0.911 | 0.917 | 0.269 | 0.923 | 0.911 | 0.916 | 0.270 | 0.920 | 0.909 | 0.914 | 0.271 | 0.920 |
| dbCAN_2:DIAMOND | 0.938 | 0.943 | 0.229 | 0.948 | 0.937 | 0.940 | 0.234 | 0.944 | 0.932 | 0.937 | 0.240 | 0.943 |
| dbCAN_2:Hotpep | 0.854 | 0.861 | 0.328 | 0.869 | 0.853 | 0.858 | 0.334 | 0.864 | 0.848 | 0.855 | 0.339 | 0.863 |
| dbCAN_3 | 0.950 | 0.955 | 0.204 | 0.959 | 0.951 | 0.954 | 0.206 | 0.957 | 0.949 | 0.954 | 0.208 | 0.958 |
| dbCAN_3:HMMER | 0.914 | 0.920 | 0.265 | 0.925 | 0.917 | 0.921 | 0.261 | 0.925 | 0.917 | 0.923 | 0.257 | 0.929 |
| dbCAN_3:DIAMOND | 0.959 | 0.963 | 0.187 | 0.967 | 0.962 | 0.965 | 0.183 | 0.967 | 0.963 | 0.966 | 0.179 | 0.970 |
| dbCAN_3:eCAMI | 0.900 | 0.907 | 0.278 | 0.913 | 0.890 | 0.894 | 0.298 | 0.899 | 0.875 | 0.882 | 0.316 | 0.889 |
| dbCAN_4 | 0.961 | 0.965 | 0.183 | 0.969 | 0.957 | 0.960 | 0.193 | 0.963 | 0.950 | 0.955 | 0.202 | 0.959 |
| dbCAN_4:HMMER | 0.919 | 0.925 | 0.259 | 0.930 | 0.920 | 0.924 | 0.258 | 0.928 | 0.918 | 0.924 | 0.256 | 0.930 |
| dbCAN_4:DIAMOND | 0.957 | 0.961 | 0.190 | 0.965 | 0.963 | 0.966 | 0.180 | 0.969 | 0.967 | 0.970 | 0.168 | 0.974 |
| dbCAN_4:dbCAN-sub | 0.963 | 0.967 | 0.177 | 0.971 | 0.955 | 0.958 | 0.195 | 0.961 | 0.945 | 0.950 | 0.212 | 0.955 |
| CUPP | 0.888 | 0.895 | 0.290 | 0.901 | 0.888 | 0.893 | 0.295 | 0.897 | 0.885 | 0.891 | 0.300 | 0.898 |

## SI table 109: Two-way ANOVA of the mean adjusted rand index between CAZyme classifiers and the taxonomic kingdom of the test set

| | Degrees of freedom | Sum of squares | Mean of squares | F value | P-value |
|---|---|---|---|---|---|
| Taxonomic group | 2 | 0 | 0.208 | 3.383 | 0.03394 |
| CAZyme classifier | 12 | 340 | 28.339 | 461.476 | <2E-16 |
| Taxonomic group : CAZyme classifier | 24 | 3 | 0.120 | 1.962 | 0.00327 |
| Residuals | 415961 | 2554 | 0.061 | | |

## SI table 110: Output of Tukey HSD test for statistically significant differences between the mean F1-score between CAZyme classifiers classifying CAZyme families from the same CAZyme class, reporting only tests with a p-value <0.05 (Overleaf)

Reporting only hits where a statistically significant difference (p-value<0.05) is found between the mean F1-scores achieved by two different classifiers for the same class (i.e. Classifier 1 and Classifier 2 are different but Class 1 and Class 2 are the same). The complete dataset is available in the online repository. Here the pair of CAZy classes, pair of classifiers/tools, the mean difference (Mean Diff), lower and upper 95% confidence interval and the adjusted p-value (Adj.p-value) are reported.

| Group 1 | Group 2 | Classifier 1 | Classifier 2 | Mean Difference | Lower 95% Confidence Interval | Upper 95% Confidence Interval | Adjusted P-value |
|---|---|---|---|---|---|---|---|
| Bacteria | Bacteria | dbCAN_2:Hotpep | dbCAN_2 | -0.067476343 | -0.082661688 | -0.052290997 | 0 |
| Bacteria | Bacteria | dbCAN_3 | dbCAN_2 | 0.020111822 | 0.004926476 | 0.035297168 | 0.000197695 |
| Bacteria | Bacteria | dbCAN_3:DIAMOND | dbCAN_2 | 0.028848904 | 0.013663558 | 0.044034249 | 1.34E-10 |
| Bacteria | Bacteria | dbCAN_3:eCAMI | dbCAN_2 | -0.025175751 | -0.040361096 | -0.009990405 | 9.70E-08 |
| Bacteria | Bacteria | dbCAN_4 | dbCAN_2 | 0.026922067 | 0.011736722 | 0.042107413 | 4.71E-09 |
| Bacteria | Bacteria | dbCAN_4:DIAMOND | dbCAN_2 | 0.027915044 | 0.012729698 | 0.04310039 | 7.74E-10 |
| Bacteria | Bacteria | dbCAN_4:sub | dbCAN_2 | 0.029719525 | 0.014534179 | 0.044904871 | 2.49E-11 |
| Bacteria | Bacteria | CUPP | dbCAN_2 | -0.025641557 | -0.040826902 | -0.010456211 | 4.41E-08 |
| Bacteria | Bacteria | dbCAN_2:DIAMOND | dbCAN_2:HMMER | 0.023815179 | 0.008629833 | 0.039000524 | 8.91E-07 |
| Bacteria | Bacteria | dbCAN_2:Hotpep | dbCAN_2:HMMER | -0.052554143 | -0.067739488 | -0.037368797 | 0 |
| Bacteria | Bacteria | dbCAN_3 | dbCAN_2:HMMER | 0.035034022 | 0.019848676 | 0.050219368 | 2.74E-13 |
| Bacteria | Bacteria | dbCAN_3:DIAMOND | dbCAN_2:HMMER | 0.043771103 | 0.028585758 | 0.058956449 | 1.13E-13 |
| Bacteria | Bacteria | dbCAN_4 | dbCAN_2:HMMER | 0.041844267 | 0.026658922 | 0.057029613 | 3.88E-13 |
| Bacteria | Bacteria | dbCAN_4:DIAMOND | dbCAN_2:HMMER | 0.042837244 | 0.027651898 | 0.05802259 | 3.83E-13 |
| Bacteria | Bacteria | dbCAN_4:sub | dbCAN_2:HMMER | 0.044641725 | 0.029456379 | 0.05982707 | 0 |
| Bacteria | Bacteria | dbCAN_2:Hotpep | dbCAN_2:DIAMOND | -0.076369321 | -0.091554667 | -0.061183975 | 0 |
| Bacteria | Bacteria | dbCAN_3:HMMER | dbCAN_2:DIAMOND | -0.020611609 | -0.035796955 | -0.005426263 | 0.000100912 |
| Bacteria | Bacteria | dbCAN_3:DIAMOND | dbCAN_2:DIAMOND | 0.019955925 | 0.004770579 | 0.035141271 | 0.000242913 |
| Bacteria | Bacteria | dbCAN_3:eCAMI | dbCAN_2:DIAMOND | -0.034068729 | -0.049254075 | -0.018883383 | 3.28E-13 |
| Bacteria | Bacteria | dbCAN_4 | dbCAN_2:DIAMOND | 0.018029089 | 0.002843743 | 0.033214434 | 0.002651563 |
| Bacteria | Bacteria | dbCAN_4:HMMER | dbCAN_2:DIAMOND | -0.017267523 | -0.032452869 | -0.002082178 | 0.006269052 |
| Bacteria | Bacteria | dbCAN_4:DIAMOND | dbCAN_2:DIAMOND | 0.019022065 | 0.00383672 | 0.034207411 | 0.000802586 |
| Bacteria | Bacteria | dbCAN_4:sub | dbCAN_2:DIAMOND | 0.020826546 | 0.005641201 | 0.036011892 | 7.51E-05 |
| Bacteria | Bacteria | CUPP | dbCAN_2:DIAMOND | -0.034534535 | -0.049719881 | -0.019349189 | 2.95E-13 |
| Bacteria | Bacteria | dbCAN_3 | dbCAN_2:Hotpep | 0.087588164 | 0.072402819 | 0.10277351 | 0 |
| Bacteria | Bacteria | dbCAN_3:HMMER | dbCAN_2:Hotpep | 0.055757712 | 0.040572367 | 0.070943058 | 0 |
| Bacteria | Bacteria | dbCAN_3:DIAMOND | dbCAN_2:Hotpep | 0.096325246 | 0.0811399 | 0.111510592 | 0 |
| Bacteria | Bacteria | dbCAN_3:eCAMI | dbCAN_2:Hotpep | 0.042300592 | 0.027115246 | 0.057485938 | 3.66E-13 |
| Bacteria | Bacteria | dbCAN_4 | dbCAN_2:Hotpep | 0.09439841 | 0.079213064 | 0.109583756 | 0 |
| Bacteria | Bacteria | dbCAN_4:HMMER | dbCAN_2:Hotpep | 0.059101798 | 0.043916452 | 0.074287144 | 0 |
| Bacteria | Bacteria | dbCAN_4:DIAMOND | dbCAN_2:Hotpep | 0.095391387 | 0.080206041 | 0.110576732 | 0 |
| Bacteria | Bacteria | dbCAN_4:sub | dbCAN_2:Hotpep | 0.097195867 | 0.082010522 | 0.112381213 | 0 |
| Bacteria | Bacteria | CUPP | dbCAN_2:Hotpep | 0.041834786 | 0.02664944 | 0.057020132 | 3.86E-13 |
| Bacteria | Bacteria | dbCAN_3:HMMER | dbCAN_3 | -0.031830452 | -0.047015798 | -0.016645107 | 8.12E-13 |
| Bacteria | Bacteria | dbCAN_3:eCAMI | dbCAN_3 | -0.045287572 | -0.060472918 | -0.030102227 | 0 |
| Bacteria | Bacteria | dbCAN_4:HMMER | dbCAN_3 | -0.028486367 | -0.043671712 | -0.013301021 | 2.66E-10 |
| Bacteria | Bacteria | CUPP | dbCAN_3 | -0.045753378 | -0.060938724 | -0.030568033 | 0 |
| Bacteria | Bacteria | dbCAN_3:DIAMOND | dbCAN_3:HMMER | 0.040567534 | 0.025382188 | 0.05575288 | 3.64E-13 |
| Bacteria | Bacteria | dbCAN_4 | dbCAN_3:HMMER | 0.038640698 | 0.023455352 | 0.053826043 | 3.72E-13 |
| Bacteria | Bacteria | dbCAN_4:DIAMOND | dbCAN_3:HMMER | 0.039633674 | 0.024448329 | 0.05481902 | 3.99E-13 |
| Bacteria | Bacteria | dbCAN_4:sub | dbCAN_3:HMMER | 0.041438155 | 0.02625281 | 0.056623501 | 3.69E-13 |
| Bacteria | Bacteria | dbCAN_3:eCAMI | dbCAN_3:DIAMOND | -0.054024654 | -0.06921 | -0.038839308 | 0 |
| Bacteria | Bacteria | dbCAN_4:HMMER | dbCAN_3:DIAMOND | -0.037223448 | -0.052408794 | -0.022038103 | 3.29E-13 |
| Bacteria | Bacteria | CUPP | dbCAN_3:DIAMOND | -0.05449046 | -0.069675806 | -0.039305114 | 0 |
| Bacteria | Bacteria | dbCAN_4 | dbCAN_3:eCAMI | 0.052097818 | 0.036912472 | 0.067283164 | 0 |
| Bacteria | Bacteria | dbCAN_4:HMMER | dbCAN_3:eCAMI | 0.016801206 | 0.00161586 | 0.031986551 | 0.010351085 |
| Bacteria | Bacteria | dbCAN_4:DIAMOND | dbCAN_3:eCAMI | 0.053090795 | 0.037905449 | 0.06827614 | 0 |
| Bacteria | Bacteria | dbCAN_4:sub | dbCAN_3:eCAMI | 0.054895275 | 0.03970993 | 0.070080621 | 0 |
| Bacteria | Bacteria | dbCAN_4:HMMER | dbCAN_4 | -0.035296612 | -0.050481958 | -0.020111266 | 2.60E-13 |
| Bacteria | Bacteria | CUPP | dbCAN_4 | -0.052563624 | -0.06774897 | -0.037378278 | 0 |
| Bacteria | Bacteria | dbCAN_4:DIAMOND | dbCAN_4:HMMER | 0.036289589 | 0.021104243 | 0.051474934 | 4.16E-13 |
| Bacteria | Bacteria | dbCAN_4:sub | dbCAN_4:HMMER | 0.03809407 | 0.022908724 | 0.053279415 | 4.54E-13 |
| Bacteria | Bacteria | CUPP | dbCAN_4:HMMER | -0.017267012 | -0.032452358 | -0.002081666 | 0.006272569 |
| Bacteria | Bacteria | CUPP | dbCAN_4:DIAMOND | -0.053556601 | -0.068741946 | -0.038371255 | 0 |
| Bacteria | Bacteria | CUPP | dbCAN_4:sub | -0.055361081 | -0.070546427 | -0.040175736 | 0 |
| All | All | dbCAN_2:HMMER | dbCAN_2 | -0.011358286 | -0.022095947 | -0.000620625 | 0.021957909 |
| All | All | dbCAN_2:DIAMOND | dbCAN_2 | 0.012662937 | 0.001925276 | 0.023400598 | 0.003050164 |
| All | All | dbCAN_2:Hotpep | dbCAN_2 | -0.062436376 | -0.073174037 | -0.051698716 | 0 |
| All | All | dbCAN_3 | dbCAN_2 | 0.024844916 | 0.014107255 | 0.035582577 | 2.67E-13 |
| All | All | dbCAN_3:DIAMOND | dbCAN_2 | 0.036318082 | 0.025580421 | 0.047055743 | 0 |
| All | All | dbCAN_3:eCAMI | dbCAN_2 | -0.029939506 | -0.040677167 | -0.019201845 | 3.49E-13 |
| All | All | dbCAN_4 | dbCAN_2 | 0.027942581 | 0.01720492 | 0.038680241 | 4.25E-13 |
| All | All | dbCAN_4:DIAMOND | dbCAN_2 | 0.037727913 | 0.026990252 | 0.048465574 | 0 |
| All | All | dbCAN_4:sub | dbCAN_2 | 0.028484594 | 0.017746933 | 0.039222255 | 4.16E-13 |
| All | All | CUPP | dbCAN_2 | -0.024736865 | -0.035474526 | -0.013999204 | 2.73E-13 |
| All | All | dbCAN_2:DIAMOND | dbCAN_2:HMMER | 0.024021222 | 0.013283561 | 0.034758883 | 3.28E-13 |
| All | All | dbCAN_2:Hotpep | dbCAN_2:HMMER | -0.051078091 | -0.061815752 | -0.04034043 | 0 |
| All | All | dbCAN_3 | dbCAN_2:HMMER | 0.036203202 | 0.025465541 | 0.046940863 | 0 |
| All | All | dbCAN_3:DIAMOND | dbCAN_2:HMMER | 0.047676368 | 0.036938707 | 0.058414029 | 0 |
| All | All | dbCAN_3:eCAMI | dbCAN_2:HMMER | -0.01858122 | -0.029318881 | -0.007843559 | 1.47E-08 |
| All | All | dbCAN_4 | dbCAN_2:HMMER | 0.039300866 | 0.028563205 | 0.050038527 | 0 |
| All | All | dbCAN_4:DIAMOND | dbCAN_2:HMMER | 0.049086199 | 0.038348538 | 0.059823859 | 0 |
| All | All | dbCAN_4:sub | dbCAN_2:HMMER | 0.03984288 | 0.029105219 | 0.050580541 | 0 |
| All | All | CUPP | dbCAN_2:HMMER | -0.013378579 | -0.02411624 | -0.002640918 | 0.000910636 |
| All | All | dbCAN_2:Hotpep | dbCAN_2:DIAMOND | -0.075099313 | -0.085836974 | -0.064361652 | 0 |
| All | All | dbCAN_3 | dbCAN_2:DIAMOND | 0.012181979 | 0.001444319 | 0.02291964 | 0.006546821 |
| All | All | dbCAN_3:HMMER | dbCAN_2:DIAMOND | -0.018442408 | -0.029180069 | -0.007704747 | 2.07E-08 |
| All | All | dbCAN_3:DIAMOND | dbCAN_2:DIAMOND | 0.023655145 | 0.012917484 | 0.034392806 | 3.70E-13 |
| All | All | dbCAN_3:eCAMI | dbCAN_2:DIAMOND | -0.042602443 | -0.053340103 | -0.031864782 | 0 |
| All | All | dbCAN_4 | dbCAN_2:DIAMOND | 0.015279644 | 0.004541983 | 0.026017305 | 2.50E-05 |
| All | All | dbCAN_4:HMMER | dbCAN_2:DIAMOND | -0.016498083 | -0.027235744 | -0.005760422 | 1.90E-06 |
| All | All | dbCAN_4:DIAMOND | dbCAN_2:DIAMOND | 0.025064976 | 0.014327315 | 0.035802637 | 5.24E-13 |
| All | All | dbCAN_4:sub | dbCAN_2:DIAMOND | 0.015821658 | 0.005083997 | 0.026559319 | 8.15E-06 |
| All | All | CUPP | dbCAN_2:DIAMOND | -0.037399801 | -0.048137462 | -0.02666214 | 0 |
| All | All | dbCAN_3 | dbCAN_2:Hotpep | 0.087281293 | 0.076543632 | 0.098018954 | 0 |
| All | All | dbCAN_3:HMMER | dbCAN_2:Hotpep | 0.056656905 | 0.045919244 | 0.067394566 | 0 |
| All | All | dbCAN_3:DIAMOND | dbCAN_2:Hotpep | 0.098754459 | 0.088016798 | 0.109492119 | 0 |
| All | All | dbCAN_3:eCAMI | dbCAN_2:Hotpep | 0.032496871 | 0.02175921 | 0.043234532 | 0 |
| All | All | dbCAN_4 | dbCAN_2:Hotpep | 0.090378957 | 0.079641296 | 0.101116618 | 0 |
| All | All | dbCAN_4:HMMER | dbCAN_2:Hotpep | 0.05860123 | 0.047863569 | 0.069338891 | 0 |
| All | All | dbCAN_4:DIAMOND | dbCAN_2:Hotpep | 0.100164289 | 0.089426628 | 0.11090195 | 0 |
| All | All | dbCAN_4:sub | dbCAN_2:Hotpep | 0.090920971 | 0.08018331 | 0.101658632 | 0 |
| All | All | CUPP | dbCAN_2:Hotpep | 0.037699512 | 0.026961851 | 0.048437173 | 0 |
| All | All | dbCAN_3:HMMER | dbCAN_3 | -0.030624388 | -0.041362049 | -0.019886727 | 2.16E-13 |
| All | All | dbCAN_3:DIAMOND | dbCAN_3 | 0.011473166 | 0.000735505 | 0.022210827 | 0.018692956 |
| All | All | dbCAN_3:eCAMI | dbCAN_3 | -0.054784422 | -0.065522083 | -0.044046761 | 0 |
| All | All | dbCAN_4:HMMER | dbCAN_3 | -0.028680063 | -0.039417724 | -0.017942402 | 3.59E-13 |
| All | All | dbCAN_4:DIAMOND | dbCAN_3 | 0.012882997 | 0.002145336 | 0.023620658 | 0.002122259 |
| All | All | CUPP | dbCAN_3 | -0.049581781 | -0.060319442 | -0.03884412 | 0 |
| All | All | dbCAN_3:DIAMOND | dbCAN_3:HMMER | 0.042097554 | 0.031359893 | 0.052835215 | 0 |
| All | All | dbCAN_3:eCAMI | dbCAN_3:HMMER | -0.024160034 | -0.034897695 | -0.013422373 | 3.17E-13 |
| All | All | dbCAN_4 | dbCAN_3:HMMER | 0.033722052 | 0.022984391 | 0.044459713 | 0 |
| All | All | dbCAN_4:DIAMOND | dbCAN_3:HMMER | 0.043507385 | 0.032769724 | 0.054245045 | 0 |
| All | All | dbCAN_4:sub | dbCAN_3:HMMER | 0.034264066 | 0.023526405 | 0.045001727 | 0 |
| All | All | CUPP | dbCAN_3:HMMER | -0.018957393 | -0.029695054 | -0.008219732 | 5.76E-09 |
| All | All | dbCAN_3:eCAMI | dbCAN_3:DIAMOND | -0.066257588 | -0.076995249 | -0.055519927 | 0 |
| All | All | dbCAN_4:HMMER | dbCAN_3:DIAMOND | -0.040153229 | -0.050890889 | -0.029415568 | 0 |
| All | All | CUPP | dbCAN_3:DIAMOND | -0.061054947 | -0.071792608 | -0.050317286 | 0 |
| All | All | dbCAN_4 | dbCAN_3:eCAMI | 0.057882086 | 0.047144425 | 0.068619747 | 0 |
| All | All | dbCAN_4:HMMER | dbCAN_3:eCAMI | 0.026104359 | 0.015366698 | 0.03684202 | 3.55E-13 |
| All | All | dbCAN_4:DIAMOND | dbCAN_3:eCAMI | 0.067667419 | 0.056929758 | 0.07840508 | 0 |
| All | All | dbCAN_4:sub | dbCAN_3:eCAMI | 0.0584241 | 0.047686439 | 0.069161761 | 0 |
| All | All | dbCAN_4:HMMER | dbCAN_4 | -0.031777727 | -0.042515388 | -0.021040066 | 0 |
| All | All | CUPP | dbCAN_4 | -0.052679445 | -0.063417106 | -0.041941784 | 0 |
| All | All | dbCAN_4:DIAMOND | dbCAN_4:HMMER | 0.041563059 | 0.030825399 | 0.05230072 | 0 |
| All | All | dbCAN_4:sub | dbCAN_4:HMMER | 0.032319741 | 0.02158208 | 0.043057402 | 0 |
| All | All | CUPP | dbCAN_4:HMMER | -0.020901718 | -0.031639379 | -0.010164057 | 3.39E-11 |
| All | All | CUPP | dbCAN_4:DIAMOND | -0.062464778 | -0.073202438 | -0.051727117 | 0 |
| All | All | CUPP | dbCAN_4:sub | -0.053221459 | -0.06395912 | -0.042483798 | 0 |
| Eukaryote | Eukaryote | dbCAN_2:DIAMOND | dbCAN_2 | 0.016432895 | 0.001247549 | 0.031618241 | 0.01516346 |
| Eukaryote | Eukaryote | dbCAN_2:Hotpep | dbCAN_2 | -0.05739641 | -0.072581756 | -0.042211065 | 0 |
| Eukaryote | Eukaryote | dbCAN_3 | dbCAN_2 | 0.02957801 | 0.014392665 | 0.044763356 | 3.28E-11 |
| Eukaryote | Eukaryote | dbCAN_3:DIAMOND | dbCAN_2 | 0.043787261 | 0.028601915 | 0.058972606 | 1.04E-13 |
| Eukaryote | Eukaryote | dbCAN_3:eCAMI | dbCAN_2 | -0.034703261 | -0.049888607 | -0.019517915 | 2.88E-13 |
| Eukaryote | Eukaryote | dbCAN_4 | dbCAN_2 | 0.028963094 | 0.013777748 | 0.044148439 | 1.07E-10 |
| Eukaryote | Eukaryote | dbCAN_4:DIAMOND | dbCAN_2 | 0.047540782 | 0.032355436 | 0.062726128 | 0 |
| Eukaryote | Eukaryote | dbCAN_4:sub | dbCAN_2 | 0.027249664 | 0.012064318 | 0.042435009 | 2.62E-09 |
| Eukaryote | Eukaryote | CUPP | dbCAN_2 | -0.023832173 | -0.039017518 | -0.008646827 | 8.68E-07 |
| Eukaryote | Eukaryote | dbCAN_2:DIAMOND | dbCAN_2:HMMER | 0.024227266 | 0.00904192 | 0.039412612 | 4.61E-07 |
| Eukaryote | Eukaryote | dbCAN_2:Hotpep | dbCAN_2:HMMER | -0.049602039 | -0.064787385 | -0.034416693 | 0 |
| Eukaryote | Eukaryote | dbCAN_3 | dbCAN_2:HMMER | 0.037372382 | 0.022187036 | 0.052557727 | 3.13E-13 |
| Eukaryote | Eukaryote | dbCAN_3:DIAMOND | dbCAN_2:HMMER | 0.051581632 | 0.036396286 | 0.066766977 | 0 |
| Eukaryote | Eukaryote | dbCAN_3:eCAMI | dbCAN_2:HMMER | -0.02690889 | -0.042094236 | -0.011723544 | 4.83E-09 |
| Eukaryote | Eukaryote | dbCAN_4 | dbCAN_2:HMMER | 0.036757465 | 0.021572119 | 0.051942811 | 3.74E-13 |
| Eukaryote | Eukaryote | dbCAN_4:DIAMOND | dbCAN_2:HMMER | 0.055335153 | 0.040149807 | 0.070520499 | 0 |
| Eukaryote | Eukaryote | dbCAN_4:sub | dbCAN_2:HMMER | 0.035044035 | 0.019858689 | 0.050229381 | 2.66E-13 |
| Eukaryote | Eukaryote | CUPP | dbCAN_2:HMMER | -0.016037801 | -0.031223147 | -0.000852456 | 0.022508686 |
| Eukaryote | Eukaryote | dbCAN_2:Hotpep | dbCAN_2:DIAMOND | -0.073829305 | -0.089014651 | -0.05864396 | 0 |
| Eukaryote | Eukaryote | dbCAN_3:HMMER | dbCAN_2:DIAMOND | -0.016273208 | -0.031458554 | -0.001087862 | 0.017821167 |
| Eukaryote | Eukaryote | dbCAN_3:DIAMOND | dbCAN_2:DIAMOND | 0.027354366 | 0.01216902 | 0.042539711 | 2.16E-09 |
| Eukaryote | Eukaryote | dbCAN_3:eCAMI | dbCAN_2:DIAMOND | -0.051136156 | -0.066321502 | -0.03595081 | 0 |
| Eukaryote | Eukaryote | dbCAN_4:HMMER | dbCAN_2:DIAMOND | -0.015728643 | -0.030913989 | -0.000543298 | 0.030329437 |
| Eukaryote | Eukaryote | dbCAN_4:DIAMOND | dbCAN_2:DIAMOND | 0.031107887 | 0.015922541 | 0.046293233 | 2.04E-12 |
| Eukaryote | Eukaryote | CUPP | dbCAN_2:DIAMOND | -0.040265068 | -0.055450413 | -0.025079722 | 4.20E-13 |
| Eukaryote | Eukaryote | dbCAN_3 | dbCAN_2:Hotpep | 0.086974421 | 0.071789075 | 0.102159766 | 0 |
| Eukaryote | Eukaryote | dbCAN_3:HMMER | dbCAN_2:Hotpep | 0.057556097 | 0.042370752 | 0.072741443 | 0 |
| Eukaryote | Eukaryote | dbCAN_3:DIAMOND | dbCAN_2:Hotpep | 0.101183671 | 0.085998325 | 0.116369017 | 0 |
| Eukaryote | Eukaryote | dbCAN_3:eCAMI | dbCAN_2:Hotpep | 0.022693149 | 0.007507804 | 0.037878495 | 5.06E-06 |
| Eukaryote | Eukaryote | dbCAN_4 | dbCAN_2:Hotpep | 0.086359504 | 0.071174158 | 0.10154485 | 0 |
| Eukaryote | Eukaryote | dbCAN_4:HMMER | dbCAN_2:Hotpep | 0.058100662 | 0.042915316 | 0.073286008 | 0 |
| Eukaryote | Eukaryote | dbCAN_4:DIAMOND | dbCAN_2:Hotpep | 0.104937192 | 0.089751847 | 0.120122538 | 0 |
| Eukaryote | Eukaryote | dbCAN_4:sub | dbCAN_2:Hotpep | 0.084646074 | 0.069460728 | 0.09983142 | 0 |
| Eukaryote | Eukaryote | CUPP | dbCAN_2:Hotpep | 0.033564238 | 0.018378892 | 0.048749583 | 3.61E-13 |
| Eukaryote | Eukaryote | dbCAN_3:HMMER | dbCAN_3 | -0.029418323 | -0.044603669 | -0.014232978 | 4.47E-11 |
| Eukaryote | Eukaryote | dbCAN_3:eCAMI | dbCAN_3 | -0.064281271 | -0.079466617 | -0.049095926 | 0 |
| Eukaryote | Eukaryote | dbCAN_4:HMMER | dbCAN_3 | -0.028873759 | -0.044059104 | -0.013688413 | 1.27E-10 |
| Eukaryote | Eukaryote | dbCAN_4:DIAMOND | dbCAN_3 | 0.017962772 | 0.002777426 | 0.033148117 | 0.002863555 |
| Eukaryote | Eukaryote | CUPP | dbCAN_3 | -0.053410183 | -0.068595529 | -0.038224837 | 0 |
| Eukaryote | Eukaryote | dbCAN_3:DIAMOND | dbCAN_3:HMMER | 0.043627573 | 0.028442228 | 0.058812919 | 1.39E-13 |
| Eukaryote | Eukaryote | dbCAN_3:eCAMI | dbCAN_3:HMMER | -0.034862948 | -0.050048294 | -0.019677602 | 2.80E-13 |
| Eukaryote | Eukaryote | dbCAN_4 | dbCAN_3:HMMER | 0.028803407 | 0.013618061 | 0.043988752 | 1.46E-10 |
| Eukaryote | Eukaryote | dbCAN_4:DIAMOND | dbCAN_3:HMMER | 0.047381095 | 0.032195749 | 0.062566441 | 0 |
| Eukaryote | Eukaryote | dbCAN_4:sub | dbCAN_3:HMMER | 0.027089977 | 0.011904631 | 0.042275322 | 3.49E-09 |
| Eukaryote | Eukaryote | CUPP | dbCAN_3:HMMER | -0.02399186 | -0.039177205 | -0.008806514 | 6.73E-07 |
| Eukaryote | Eukaryote | dbCAN_3:eCAMI | dbCAN_3:DIAMOND | -0.078490522 | -0.093675867 | -0.063305176 | 0 |
| Eukaryote | Eukaryote | dbCAN_4:HMMER | dbCAN_3:DIAMOND | -0.043083009 | -0.058268355 | -0.027897663 | 2.94E-13 |
| Eukaryote | Eukaryote | dbCAN_4:sub | dbCAN_3:DIAMOND | -0.016537597 | -0.031722942 | -0.001352251 | 0.013621685 |
| Eukaryote | Eukaryote | CUPP | dbCAN_3:DIAMOND | -0.067619433 | -0.082804779 | -0.052434088 | 0 |
| Eukaryote | Eukaryote | dbCAN_4 | dbCAN_3:eCAMI | 0.063666355 | 0.048481009 | 0.078851701 | 0 |
| Eukaryote | Eukaryote | dbCAN_4:HMMER | dbCAN_3:eCAMI | 0.035407513 | 0.020222167 | 0.050592858 | 2.53E-13 |
| Eukaryote | Eukaryote | dbCAN_4:DIAMOND | dbCAN_3:eCAMI | 0.082244043 | 0.067058697 | 0.097429389 | 0 |
| Eukaryote | Eukaryote | dbCAN_4:sub | dbCAN_3:eCAMI | 0.061952925 | 0.046767579 | 0.077138271 | 0 |
| Eukaryote | Eukaryote | dbCAN_4:HMMER | dbCAN_4 | -0.028258842 | -0.043444188 | -0.013073496 | 4.08E-10 |
| Eukaryote | Eukaryote | dbCAN_4:DIAMOND | dbCAN_4 | 0.018577688 | 0.003392342 | 0.033763034 | 0.001383785 |
| Eukaryote | Eukaryote | CUPP | dbCAN_4 | -0.052795266 | -0.067980612 | -0.037609921 | 0 |
| Eukaryote | Eukaryote | dbCAN_4:DIAMOND | dbCAN_4:HMMER | 0.04683653 | 0.031651185 | 0.062021876 | 0 |
| Eukaryote | Eukaryote | dbCAN_4:sub | dbCAN_4:HMMER | 0.026545412 | 0.011360066 | 0.041730758 | 9.20E-09 |
| Eukaryote | Eukaryote | CUPP | dbCAN_4:HMMER | -0.024536424 | -0.03972177 | -0.009351079 | 2.79E-07 |
| Eukaryote | Eukaryote | dbCAN_4:sub | dbCAN_4:DIAMOND | -0.020291118 | -0.035476464 | -0.005105772 | 0.000155648 |
| Eukaryote | Eukaryote | CUPP | dbCAN_4:DIAMOND | -0.071372955 | -0.0865583 | -0.056187609 | 0 |
| Eukaryote | Eukaryote | CUPP | dbCAN_4:sub | -0.051081836 | -0.066267182 | -0.035896491 | 0 |

# 16 Evaluation of CAZy family classification per taxonomic kingdom

The performance of CAZy family classification per taxonomic kingdom is evaluated across all CAZy families in section 16.1.

The performance of CAZy family classification per CAZy class is then explored in sections 17.1 and 17.1.

In section 17.1 the mean (95% CI and standard deviation) are calculated per CAZy class, and are plotted per statistical parameter to facilitate comparing across CAZy classes.

## 16.1 CAZy family classification per taxonomic kingdom across all CAZy families

### SI figure 50: The mean F1-score and 95% confidence interval of the CAZy family classification across all CAZy classes

All test sets were pooled and the F1-score calculated across per CAZy families. The mean F1-score was then calculated across all CAZy families, as well as the 95% confidence interval (CI).

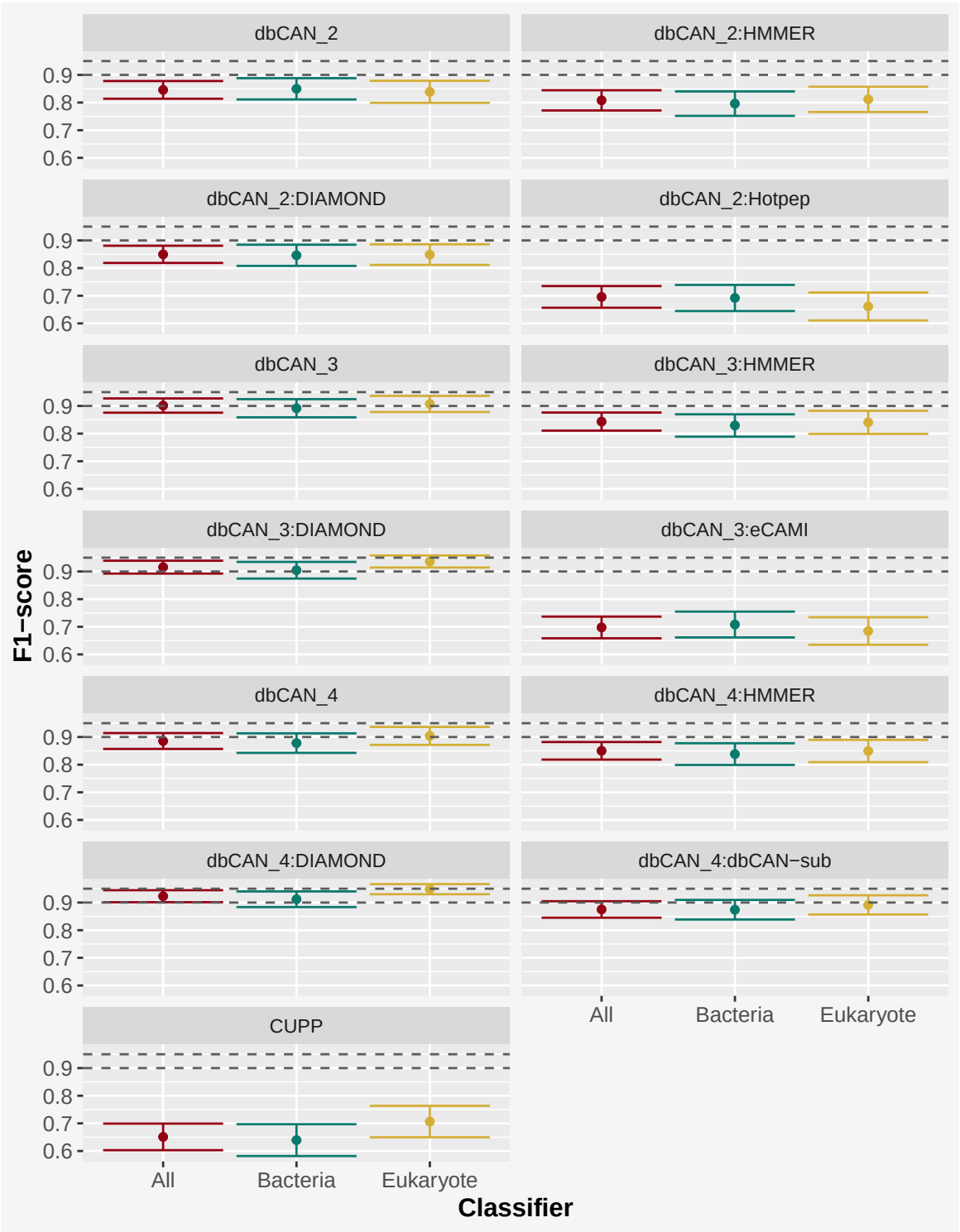



Figure 50: The mean F1-score and 95% confidence interval (CI) of binary CAZy family classification across all CAZy classes. Where a tool is incorporated into a CAZyme classifier, the tool name is prefixed with the respective CAZyme classifier. dbCAN_2 and associated tools are coloured blue; dbCAN_3 pink; dbCAN_4 green; and CUPP orange.

**SI figure 51: The F1-score per taxonomic kingdom of the CAZy family classification across all CAZy classes**

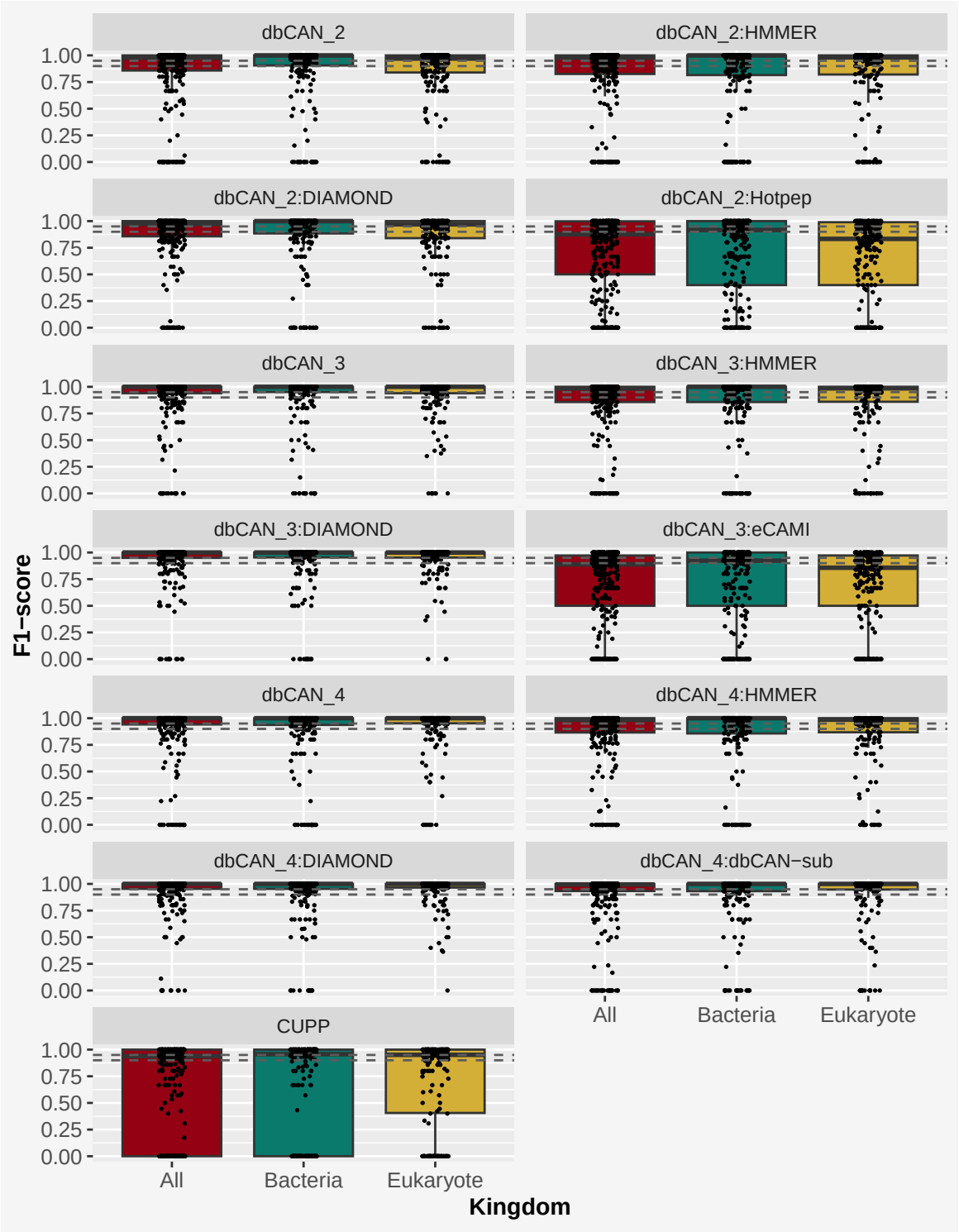



Figure 51: One-dimensional scatter plot overlaying a box and whisker plot, where each point represents the F1-score of the binary CAZy family classification of a CAZy family. Where a tool is incorporated into a CAZyme classifier, the tool name is prefixed with the respective CAZyme classifier. dbCAN_2 and associated tools are coloured blue; dbCAN_3 pink; dbCAN_4 green; and CUPP orange.

## SI figure 52: The mean sensitivity and 95% confidence interval of the CAZy family classification across all CAZy classes

All test sets were pooled and the sensitivity calculated across per CAZy families. The mean sensitivity was then calculated across all CAZy families, as well as the 95% confidence interval (CI).

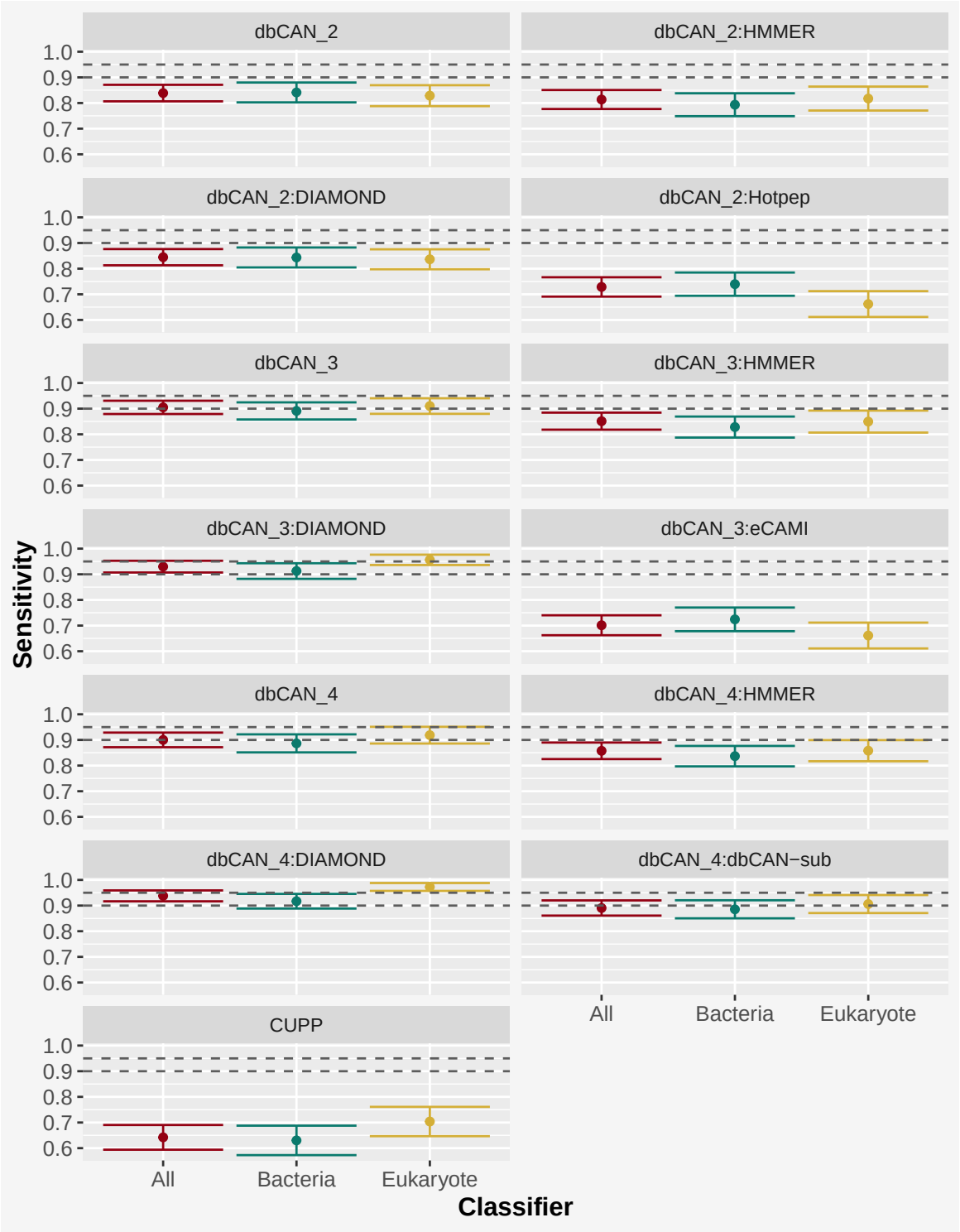



Figure 52: The mean sensitivity and 95% confidence interval (CI) of binary CAZy family classification across all CAZy classes. Where a tool is incorporated into a CAZyme classifier, the tool name is prefixed with the respective CAZyme classifier. dbCAN_2 and associated tools are coloured blue; dbCAN_3 pink; dbCAN_4 green; and CUPP orange.

**SI figure 53: The sensitivity per taxonomic kingdom of the CAZy family classification across all CAZy classes**

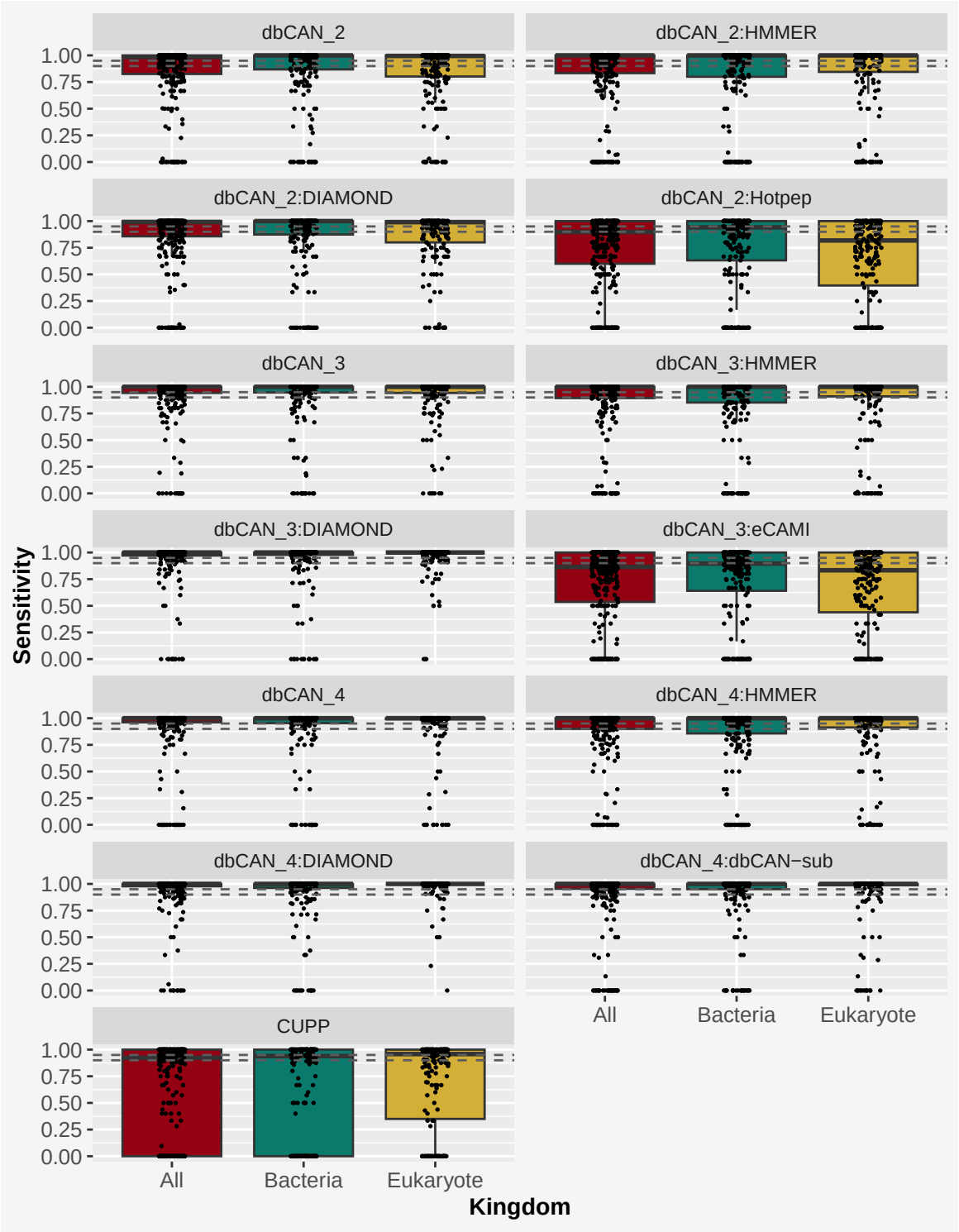



Figure 53: One-dimensional scatter plot overlaying a box and whisker plot, where each point represents the sensitivity of the binary CAZy family classification of a CAZy family. Where a tool is incorporated into a CAZyme classifier, the tool name is prefixed with the respective CAZyme classifier. dbCAN_2 and associated tools are coloured blue; dbCAN_3 pink; dbCAN_4 green; and CUPP orange.

**SI table 111: Output of a two-way ANOVA testing for statistically significant differences in the F1-score of CAZyme family classification across bacterial, eukaryotic, or all sequences.**

| | Degrees of freedom | Sum of squares | Mean of squares | F value | P-value |
|---|---|---|---|---|---|
| Tax Group | 2 | 0.217843 | 0.108921 | 1.150539 | 0.316507 |
| Prediction Tool | 12 | 75.29954 | 6.274961 | 66.28262 | 3.08E-156 |
| Tax group:Prediction tool | 24 | 1.092813 | 0.045534 | 0.480976 | 0.984524 |
| Residuals | 10132 | 959.1942 | 0.09467 | NA | NA |

**SI table 112: Output of a two-way ANOVA testing for statistically significant differences in the sensitivity of CAZyme family classification across bacterial, eukaryotic, or all sequences.**

| | Degrees of freedom | Sum of squares | Mean of squares | F value | P-value |
|---|---|---|---|---|---|
| Tax Group | 2 | 0.195998 | 0.097999 | 1.032059 | 0.35631 |
| Prediction Tool | 12 | 78.00143 | 6.500119 | 68.45492 | 2.05E-161 |
| Tax group:Prediction tool | 24 | 2.855504 | 0.118979 | 1.253011 | 0.182761 |
| Residuals | 10132 | 962.0814 | 0.094955 | NA | NA |

# 17 Testing for statistically significant differences in the mean F1-score for CAZyme family classification between the taxonomic groups, per CAZyme class

**SI table 113: Output of a two-way ANOVA testing for statistically significant differences in the sensitivity of PL CAZyme family classification across bacterial, eukaryotic, or all sequences.**

| | Degrees of freedom | Sum of squares | Mean of squares | F value | P-value |
|---|---|---|---|---|---|
| Tax Group | 2 | 0.012401 | 0.006201 | 0.071071 | 0.931403 |
| Prediction Tool | 12 | 8.450375 | 0.704198 | 8.071568 | 3.18E-14 |
| Tax group:Prediction tool | 24 | 0.448787 | 0.018699 | 0.214335 | 0.999981 |
| Residuals | 652 | 56.88325 | 0.087244 | NA | NA |

**SI table 114: Output of a two-way ANOVA testing for statistically significant differences in the sensitivity of AA CAZyme family classification across bacterial, eukaryotic, or all sequences.**

| | Degrees of freedom | Sum of squares | Mean of squares | F value | P-value |
|---|---|---|---|---|---|
| Tax Group | 2 | 0.294648 | 0.147324 | 1.713543 | 0.181627 |
| Prediction Tool | 12 | 4.37506 | 0.364588 | 4.240572 | 2.71E-06 |
| Tax group:Prediction tool | 24 | 0.127717 | 0.005322 | 0.061895 | 1 |
| Residuals | 377 | 32.41303 | 0.085976 | NA | NA |

## 17.1 CAZy family classification per taxonomic kingdom per statistical parameter

### SI figure 54: The mean F1-score and 95% confidence interval of the CAZy family classification per CAZy classes

All test sets were pooled and the F1-score calculated across per CAZy families. The mean F1-score was then calculated across all CAZy families, as well as the 95% confidence interval (CI).

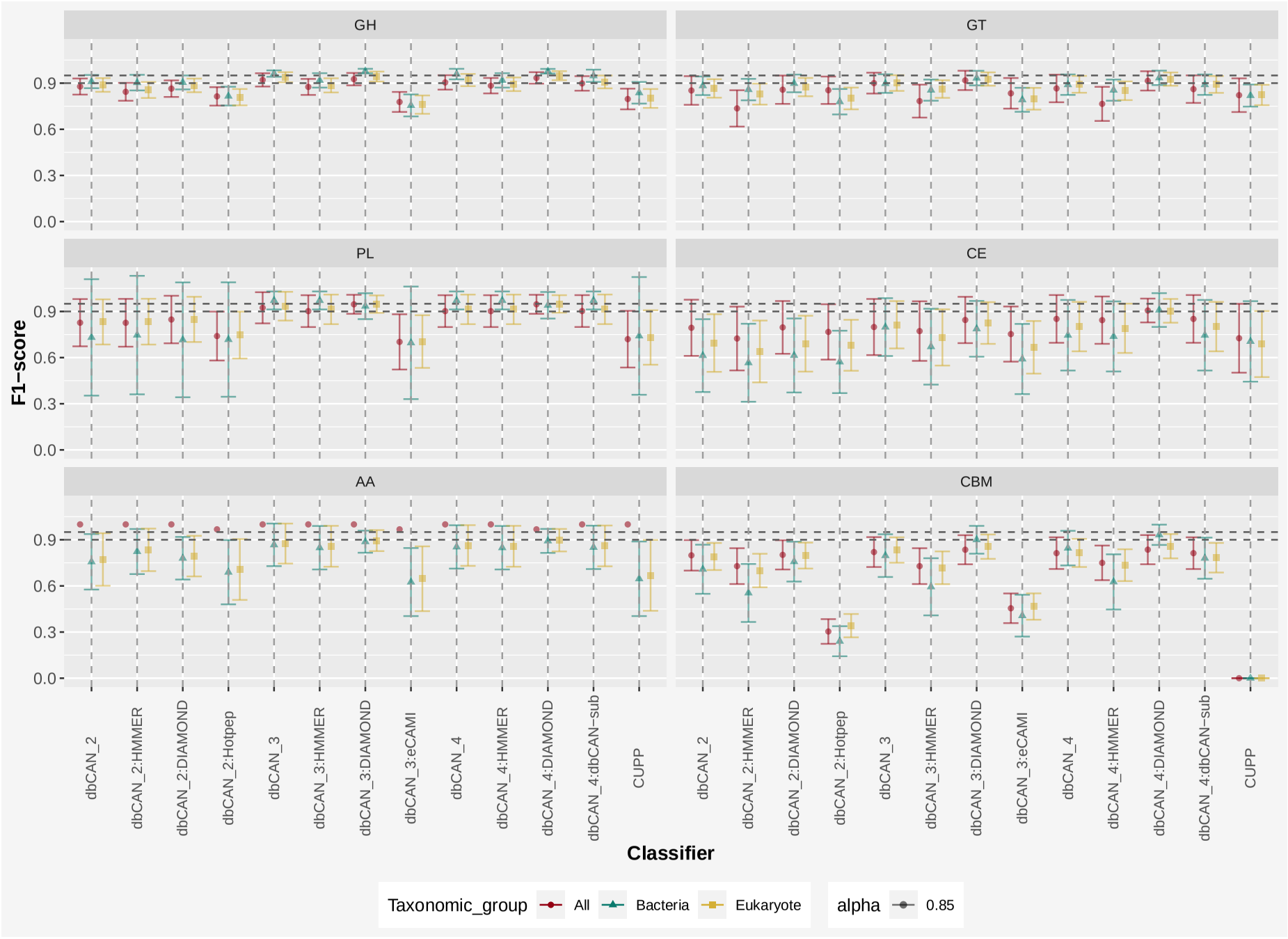



Figure 54: The mean F1-score and 95% confidence interval (CI) of binary CAZy family classification across all CAZy classes. Where a tool is incorporated into a CAZyme classifier, the tool name is prefixed with the respective CAZyme classifier. dbCAN_2 and associated tools are coloured blue; dbCAN_3 pink; dbCAN_4 green; and CUPP orange.

**SI figure 55: The F1-score per taxonomic kingdom of the CAZy family classification per CAZy classes**

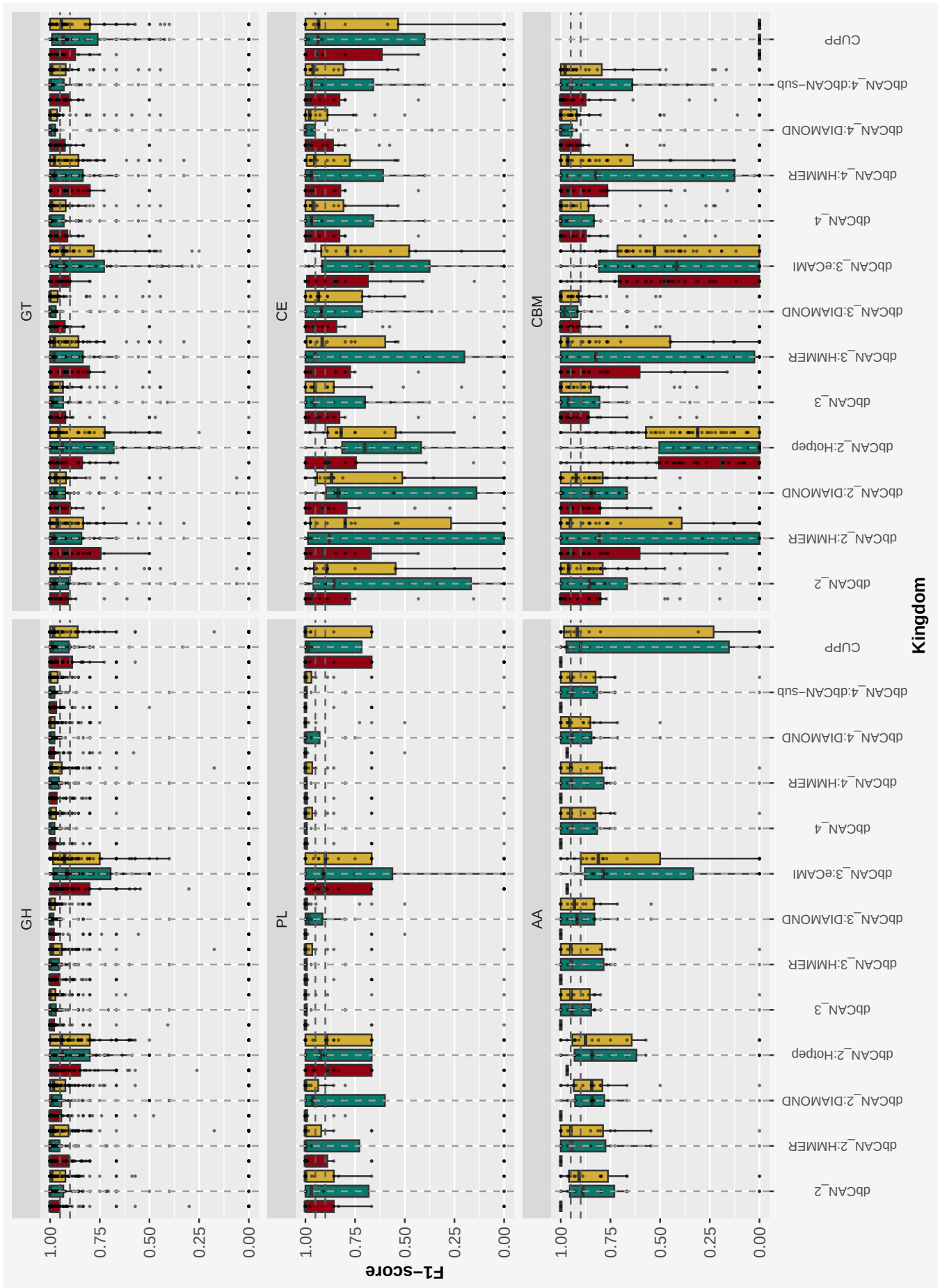



Figure 55: One-dimensional scatter plot overlaying a box and whisker plot, where each point represents the F1-score of the binary CAZy family classification of a CAZy family. Where a tool is incorporated into a CAZyme classifier, the tool name is prefixed with the respective CAZyme classifier. dbCAN_2 and associated tools are coloured blue; dbCAN_3 pink; dbCAN_4 green; and CUPP orange.

## SI figure 56: The mean sensitivity and 95% confidence interval of the CAZy family classification per CAZy classes

All test sets were pooled and the sensitivity calculated across per CAZy families. The mean sensitivity was then calculated across all CAZy families, as well as the 95% confidence interval (CI).

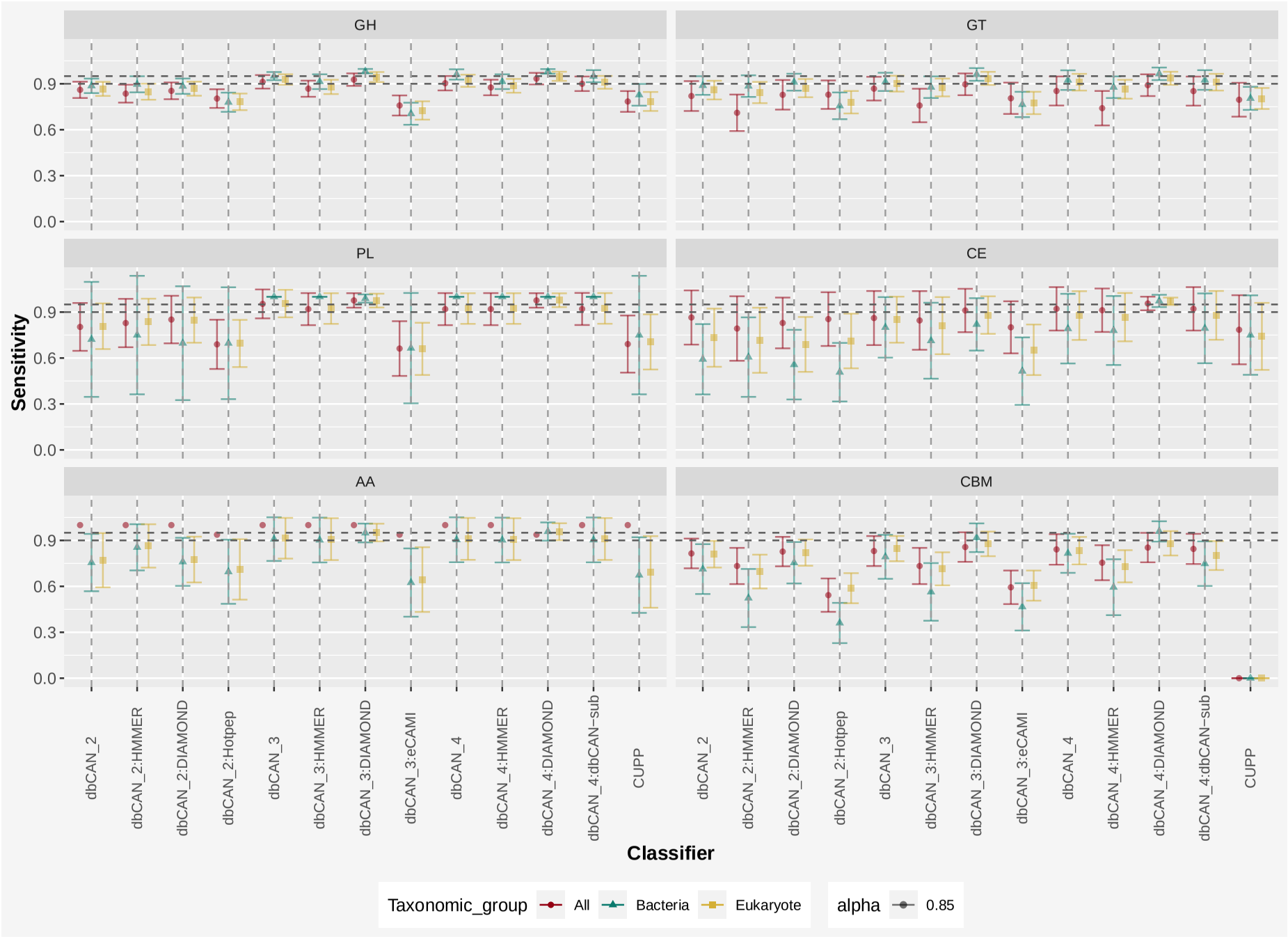



Figure 56: The mean sensitivity and 95% confidence interval (CI) of binary CAZy family classification across all CAZy classes. Where a tool is incorporated into a CAZyme classifier, the tool name is prefixed with the respective CAZyme classifier. dbCAN_2 and associated tools are coloured blue; dbCAN_3 pink; dbCAN_4 green; and CUPP orange.

**SI figure 57: The sensitivity per taxonomic kingdom of the CAZy family classification per CAZy classes**

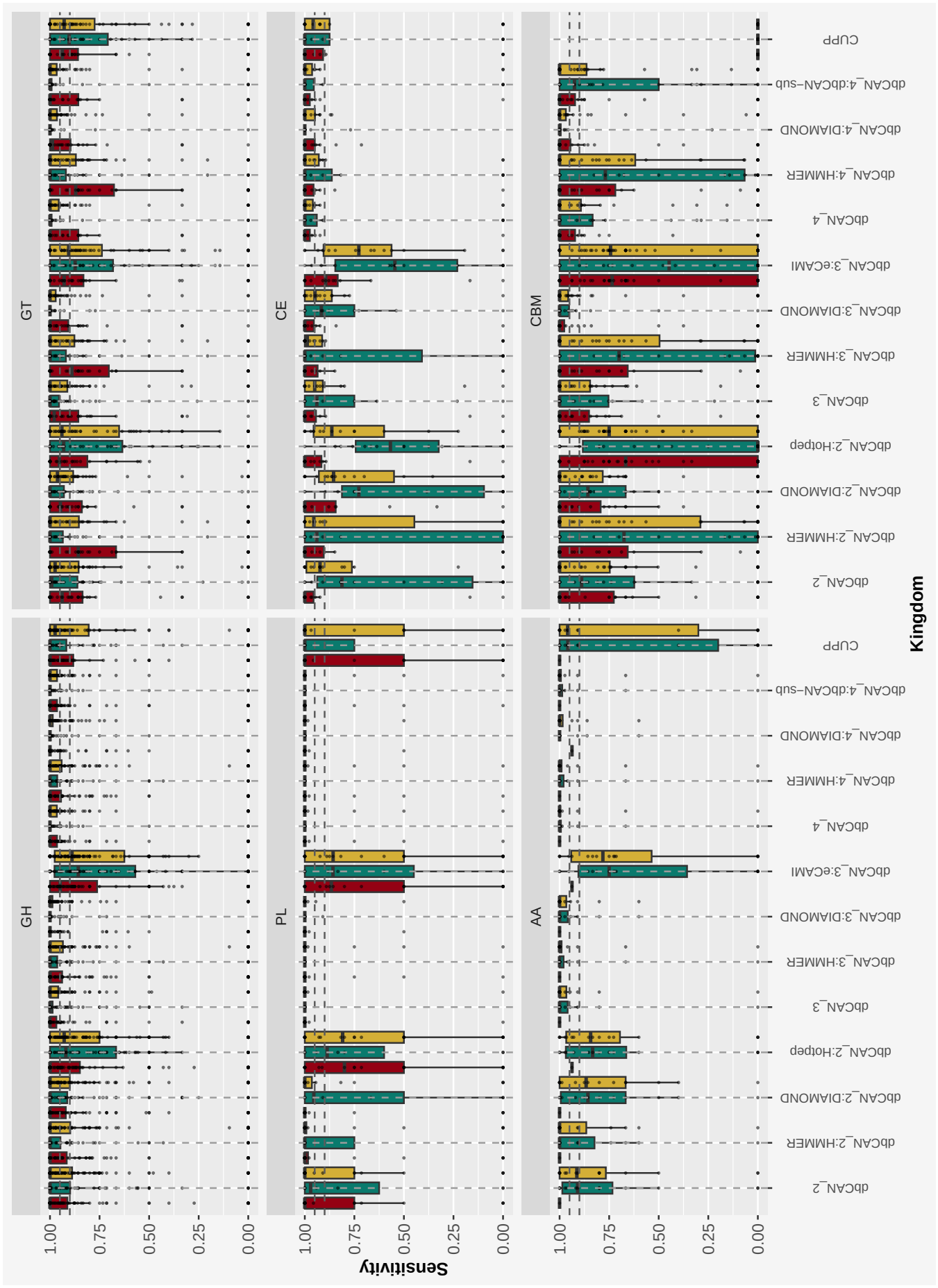



Figure 57: One-dimensional scatter plot overlaying a box and whisker plot, where each point represents the sensitivity of the binary CAZy family classification of a CAZy family. Where a tool is incorporated into a CAZyme classifier, the tool name is prefixed with the respective CAZyme classifier. dbCAN_2 and associated tools are coloured blue; dbCAN_3 pink; dbCAN_4 green; and CUPP orange.

# 18 SI table 115: Test sets with consistently poor performing CAZyme class classification (measured by an F1-score of less than 0.75)

Test sets with F1-scores of less than 0.75 from at least one classifier (including tools incorporated into dbCAN). Classifiers that achieve an F1-score of less than 0.75 are marked with 'Y'. The genomic version accession of test sets where seven or more classifiers produce an F1-score of less than 0.75 are shaded dark grey. The genomic accession of test sets where an F1-score of less than 0.75 is achieved for two catalytic CAZy classes are shaded blue; three CAZy classes are shaded green; four CAZy classes yellow; all five catalytic CAZy classes are shaded orange. The total number of test sets with an F1-score of less than or equal to 0.75 is counted per CAZy class and across all CAZy classes, and these values overlay bar charts that are relative for values per CAZy class.

| Kingdom | Organism | Genomic Accession | CAZy Class | dbCAN_2 | dbCAN_2: HMMER | dbCAN_2: DIAMOND | dbCAN_2: Hotpep | dbCAN_3 | dbCAN_3: HMMER | dbCAN_3: DIAMOND | dbCAN_3: eCAMI | dbCAN_4 | dbCAN_4: HMMER | dbCAN_4: DIAMOND | dbCAN_4: dbCAN-sub | CUPP | #ofTools |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Bacteria | *Salmonella enterica subsp. Arizonae* | GCA_900635675.1 | GT | 1 | 1 | 1 | | | 1 | 1 | | 1 | 1 | 1 | | | 8 |
| | | | PL | 1 | 1 | 1 | 1 | | 1 | | 1 | 1 | 1 | | 1 | 1 | 10 |
| | | | GH | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 13 |
| | *Actinobacillus equuli* | GCA_900638075.1 | GT | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 13 |
| | | | CE | 1 | 1 | 1 | | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | | 11 |
| | *Streptomyces antimycoticus* | GCA_009936315.1 | GT | | | | 1 | | | | | | | | | | 1 |
| | | | PL | | | | | | | | 1 | | | | | 1 | 2 |
| | | | CE | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 13 |
| | *Schleiferilactobacillus harbinensis* | GCA_008694105.1 | GT | 1 | | 1 | 1 | | | | 1 | | | | | | 4 |
| | | | PL | 1 | 1 | 1 | 1 | | 1 | | 1 | 1 | 1 | | 1 | 1 | 10 |
| | | | CE | | | 1 | | | | | | | | | | | 1 |
| | *Streptosporangium roseum DSM 43021* | GCA_000024865.1 | GT | | | | 1 | | | | | | | | | | 1 |
| | | | PL | 1 | 1 | | 1 | 1 | 1 | | 1 | 1 | 1 | | 1 | 1 | 10 |
| | | | AA | | | | | | | | | | | 1 | | | 1 |
| | *Cellvibrio japonicus* | GCA_000019225.1 | GT | | 1 | | 1 | | 1 | | 1 | | 1 | | | | 5 |
| | | | CE | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | | | | | 1 | 9 |
| | | | AA | | | | 1 | | | | 1 | | | | | | 2 |
| | *Azospirillum brasilense* | GCA_001315015.1 | GT | | | | 1 | | | | | | | | | | 1 |
| | | | PL | | | | | | | | 1 | | | | | | 1 |
| | | | CE | | 1 | 1 | | | 1 | 1 | 1 | | 1 | 1 | | | 7 |
| | *Alicyclobacillus sp. SO9* | GCA_016406125.1 | GH | | | 1 | 1 | | | | 1 | | | | | | 3 |
| | | | GT | 1 | | 1 | 1 | | | | 1 | | | | | | 4 |
| | | | CE | | | 1 | | | | | 1 | | | | | | 2 |
| | *Nibricoccus aquaticus* | GCA_002310495.1 | GT | | | | 1 | | | | 1 | | | | | 1 | 3 |
| | | | PL | | | | 1 | | | | 1 | | | | | | 2 |
| | | | CE | | | | 1 | | | | 1 | | | | | 1 | 3 |
| | *Caulobacter segnis* | GCA_000092285.1 | GT | | | | 1 | | | | | | | | | | 1 |
| | | | CE | | 1 | | | | 1 | | 1 | | | | | | 3 |
| | | | GT | | | | 1 | | | | | | | | | | 1 |
| | *Nocardia arthritidis* | GCA_011801145.1 | GH | | | | 1 | | | | | | | | | | 1 |
| | | | GT | | | | 1 | | | | | | | | | | 1 |
| | | | CE | | | 1 | | | | | | | | | | | 1 |
| | *Pseudobacter ginsenosidimutans* | GCA_007970185.1 | GT | 1 | | 1 | 1 | | | | 1 | | | | | 1 | 5 |
| | | | CE | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | | | | | 1 | 9 |
| | *Klebsiella oxytoca* | GCA_002906395.1 | PL | 1 | 1 | 1 | 1 | | | | 1 | | | | | 1 | 6 |
| | | | CE | | 1 | | | | 1 | | | | 1 | | | | 3 |
| | *Enterococcus casseliflavus* | GCA_009707345.1 | GT | | 1 | | 1 | | 1 | | | | 1 | | | | 4 |
| | | | PL | | | | | | | | | | | | | 1 | 1 |
| | *Dictyoglomus turgidum* | GCA_000021645.1 | GT | | | | 1 | | | | 1 | | | | | 1 | 3 |
| | | | PL | | | | 1 | | | | 1 | | | | | | 2 |
| | *Clostridium saccharoperbutylacetonicum* | GCA_002003305.1 | GT | | | | 1 | | | | | | | | | 1 | 2 |
| | | | CE | | | | 1 | | | | | | | | | | 1 |
| | *Streptacidiphilus sp. P02-A3a* | GCA_014084105.1 | PL | | | | 1 | | | | | | | | | | 1 |
| | | | CE | | | 1 | | | | | | | | | | | 1 |
| | *Acetivibrio clariflavus* | GCA_000237085.1 | GT | | | | 1 | | | | | | | | | | 1 |
| | | | PL | | | | | | | | 1 | | | | | | 1 |
| | *Mycobacterium sp. JS623* | GCA_000328565.1 | GH | | | | 1 | | | | | | | | | | 1 |
| | | | GT | | | | 1 | | | | | | | | | | 1 |
| | *Xanthomonas citri* | GCA_000349225.1 | CE | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | | | | | 1 | 9 |
| | *Xanthomonas citri* | GCA_000816885.1 | | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | | | | | 1 | 9 |
| | *Klebsiella michiganensis* | GCA_010093005.1 | PL | 1 | 1 | 1 | 1 | | | | 1 | | | | | 1 | 6 |
| | *Ruminiclostridium cellulolyticum* | GCA_000022065.1 | | | | | | | | | 1 | | | | | | 1 |
| | *Clostridium beijerinckii* | GCA_000833105.2 | GT | | 1 | | 1 | | 1 | | | | 1 | | | | 4 |
| | *Acetivibrio thermocellus* | GCA_000015865.1 | | | | | 1 | | | | 1 | | | | | | 2 |
| | *Paenibacillus polymyxa* | GCA_000507205.2 | | | | | 1 | | | | | | | | | | 1 |
| Eukaryote | *Ceratobasidium sp. AG-Ba* | GCA_016906575.1 | GH | | | | 1 | | | | 1 | | | | | | 2 |
| | | | GT | 1 | 1 | 1 | 1 | 1 | 1 | | 1 | | 1 | | | 1 | 9 |
| | | | PL | 1 | 1 | 1 | 1 | | | | 1 | | | | | 1 | 6 |
| | | | CE | 1 | | 1 | 1 | | | | 1 | | | | | 1 | 5 |
| | | | AA | 1 | | 1 | 1 | | | | 1 | | | | | | 4 |
| | *Trichoderma asperellum* | GCA_020647865.1 | GH | | | | 1 | | | | | | | | | | 1 |
| | | | GT | | | | 1 | | | | | | | | | 1 | 2 |
| | | | PL | 1 | 1 | 1 | 1 | | | | 1 | | | | | 1 | 6 |
| | | | CE | 1 | | 1 | 1 | 1 | | 1 | 1 | | | | | | 6 |
| | | | AA | 1 | | | 1 | | | | 1 | | | | | 1 | 4 |
| | *Trichoderma atroviride* | GCA_020647795.1 | GT | 1 | 1 | | 1 | | 1 | | 1 | | 1 | | | 1 | 7 |
| | | | PL | 1 | | 1 | 1 | | | | 1 | | | | | | 4 |
| | | | CE | | | 1 | 1 | 1 | | 1 | 1 | | | | | | 5 |
| | | | AA | | | | | | | | 1 | | | | | | 1 |
| | *Fusariumsolani-melongenae* | GCA_023101225.1 | GT | | | | 1 | | | | | | | | | | 1 |
| | | | PL | 1 | 1 | 1 | 1 | | | 1 | 1 | | | | | 1 | 7 |
| | | | CE | | | | 1 | | | | 1 | | | | | | 2 |
| | | | AA | | | | | | | | 1 | | | | | | 1 |
| | *Pyricularia oryzae* | GCA_004346965.1 | GT | | | | | | | | 1 | | | | | | 1 |
| | | | PL | 1 | 1 | 1 | 1 | | | | 1 | | | | | 1 | 6 |
| | | | CE | | | | | | | | 1 | | | | | | 1 |
| | | | AA | | 1 | | | | | | | | | | | 1 | 2 |
| | *Fulvia fulva* | GCA_020509005.2 | GH | | | | 1 | | | | | | | | | | 1 |
| | | | GT | | 1 | | | | 1 | | | | 1 | | | 1 | 4 |
| | | | CE | | | 1 | | 1 | | 1 | 1 | | | | | | 4 |
| | | | AA | | | | 1 | | | | 1 | | | | | | 2 |
| | *Verrucomicrobia bacterium* | GCA_000972765.1 | GH | | | | 1 | | | | 1 | | | | | | 2 |
| | | | GT | | | | 1 | | | | 1 | | | | | | 2 |
| | | | PL | | | | 1 | | | | 1 | | | | | 1 | 3 |
| | | | CE | | | | | | | | 1 | | | | | 1 | 2 |
| | *Botrytis cinerea* | GCA_000143535.4 | GT | | | | 1 | | | | | | | | | 1 | 2 |
| | | | PL | | | | 1 | | | | 1 | | | | | | 2 |
| | | | CE | | | | | | | | 1 | | | | | 1 | 2 |
| | | | AA | | | | 1 | | | | 1 | | | | | | 2 |
| | *Micromonas commoda* | GCA_000090985.2 | GH | | | | 1 | | | | | | | | | | 1 |
| | | | GT | | | | 1 | | | | 1 | | | | | | 2 |
| | | | CE | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | | | | | 1 | 9 |
| | *Chloropicon primus* | GCA_007859695.1 | GH | 1 | | 1 | 1 | | | | 1 | | | | | | 4 |
| | | | GT | 1 | | 1 | 1 | | | | 1 | | | | | 1 | 5 |
| | | | AA | | | | | | | | 1 | | | | | | 1 |
| | *Trichoderma simmonsii* | GCA_019565615.1 | GT | | | | 1 | | | | | | | | | 1 | 2 |
| | | | CE | 1 | | 1 | 1 | | | | 1 | | | | | | 4 |
| | | | AA | 1 | | 1 | 1 | | | | 1 | | | | | 1 | 5 |
| | *Ostreococcus lucimarinus* | GCA_000092065.1 | GH | | | | 1 | | | | 1 | | | | | | 2 |
| | | | GT | | | | 1 | | | | 1 | | | | | 1 | 3 |
| | | | CE | | | | 1 | | | | 1 | | | | | 1 | 3 |
| | *Thermothelomyces thermophilus* | GCA_000226095.1 | GT | | 1 | | | | 1 | | | | 1 | | | | 3 |
| | | | CE | | | | 1 | | | | 1 | | | | | 1 | 3 |
| | | | AA | | | | | | | | | | | | | 1 | 1 |
| | *Sugiyamaella lignohabitans* | GCA_001640025.2 | GH | | | | | | | | 1 | | | | | | 1 |
| | | | CE | | | | 1 | | | | 1 | | | | | | 2 |
| | | | AA | | | 1 | | | | | | | | 1 | | | 2 |
| | *Aspergillus flavus* | GCA_009017415.1 | PL | | | | | | | | 1 | | | | | | 1 |
| | | | CE | | | | 1 | | | | 1 | | | | | | 2 |
| | | | AA | | 1 | | | | | | | | | | | 1 | 2 |
| | *Aspergillus flavus* | GCA_014784225.2 | GT | | 1 | | 1 | | 1 | | | | 1 | | | 1 | 5 |
| | | | CE | 1 | 1 | 1 | 1 | | | | 1 | | | | | | 5 |
| | *Peltaster fructicola* | GCA_001592805.2 | GT | 1 | | | 1 | | | | | | | | | 1 | 3 |
| | | | CE | 1 | | 1 | 1 | | | | 1 | | | | | | 4 |
| | *Caenorhabditis elegans* | GCA_000002985.3 | GH | | | | 1 | | | | 1 | | | | | 1 | 3 |
| | | | GT | | | | 1 | | | | | | | | | | 1 |
| | *Pichia kudriavzevii* | GCA_003054445.1 | GT | | | | 1 | | | | 1 | | | | | | 2 |
| | | | AA | | | | | | | | 1 | | | | | | 1 |
| | *Cordyceps militaris* | GCA_008080495.1 | GH | | | | 1 | | | | | | | | | | 1 |
| | | | CE | | | | 1 | | | | 1 | | | | | | 2 |
| | *Nakaseomyces glabratus* | GCA_014217725.1 | CE | | | 1 | | | | | | | | | | | 2 |
| | | | AA | | | 1 | | | | | | | | | | | 1 |
| | *Metarhizium brunneum* | GCA_013426205.1 | GT | | | | 1 | | | | | | | | | | 1 |
| | | | PL | | | 1 | | | | | | | | | | | 1 |
| | *Fusarium culmorum* | GCA_016952355.1 | GT | | | | 1 | | | | | | | | | | 1 |
| | | | CE | | | | 1 | | | | | | | | | | 1 |
| | *Fusarium pseudograminearum* | GCA_016952305.1 | GT | | | | 1 | | | | | | | | | | 1 |
| | | | CE | | | | 1 | | | | | | | | | | 1 |
| | *Saccharomyces uvarum* | GCA_027557585.1 | AA | 1 | | 1 | 1 | | | | 1 | | | | | 1 | 5 |
| | *Yarrowia lipolytica* | GCA_014490615.1 | | 1 | | 1 | 1 | | | | 1 | | | | | 1 | 5 |
| | *Clavispora lusitaniae* | GCA_009498115.1 | | 1 | | 1 | 1 | | | | 1 | | | | | 1 | 5 |
| | *[Candida] auris* | GCA_003013715.2 | | 1 | | 1 | 1 | | | | 1 | | | | | 1 | 5 |
| | *Brettanomyces nanus* | GCA_011074865.2 | | | | 1 | | | | | 1 | | | | | | 2 |
| | *Brettanomyces bruxellensis* | GCA_011074885.2 | | | | 1 | | | | | 1 | | | | | | 2 |
| | *Metschnikowia aff. Pulcherrima* | GCA_004217705.1 | | | | 1 | | | | | 1 | | | | | | 2 |
| | *Zygosaccharomyces parabailii* | GCA_001984395.2 | | | 1 | | | | | | | | | 1 | | | 2 |
| | *Kluyveromyces marxianus* | GCA_001854445.2 | | | 1 | | | | | | | | | | | | 1 |
| | *[Candida] auris* | GCA_008275145.1 | | | | 1 | | | | | | | | | | | 1 |
| | *Penicillium digitatum* | GCA_016767815.1 | CE | | | | 1 | | | | 1 | | | | | | 2 |
| | *Fusarium oxysporum* | GCA_013085055.1 | | | | | 1 | | | | | | | | | | 1 |
| | *Aspergillus chevalieri* | GCA_016861735.1 | | | | 1 | | | | | | | | | | | 1 |
| | | | **GH** | 2 | 1 | 3 | 13 | 1 | 1 | 1 | 8 | 1 | 1 | 1 | 1 | 2 | 36 |
| | | | **GT** | 9 | 10 | 7 | 37 | 2 | 10 | 2 | 16 | 2 | 10 | 2 | 1 | 15 | 123 |
| | | | **PL** | 10 | 9 | 11 | 15 | 1 | 3 | 1 | 19 | 3 | 3 | 0 | 3 | 12 | 89 |
| | | | **CE** | 32 | 11 | 21 | 24 | 10 | 10 | 11 | 29 | 2 | 4 | 3 | 2 | 12 | 151 |
| | | | **AA** | 4 | 2 | 5 | 7 | 0 | 0 | 0 | 11 | 0 | 0 | 2 | 0 | 6 | 37 |
| | | | **Total** | 37 | 33 | 47 | 96 | 14 | 24 | 15 | 82 | 8 | 18 | 8 | 7 | 47 | 436 |