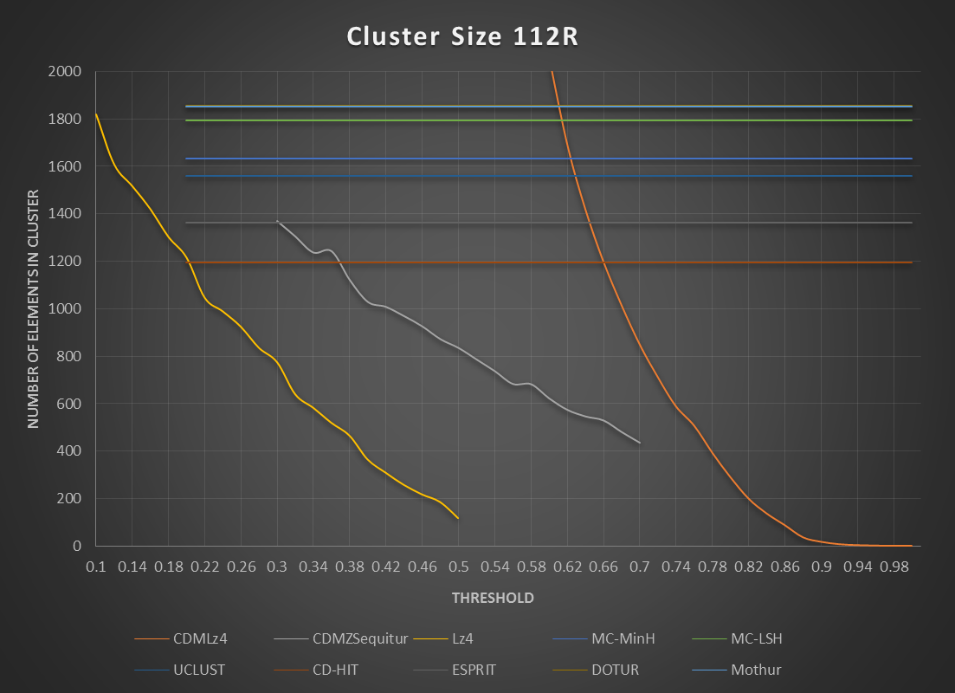
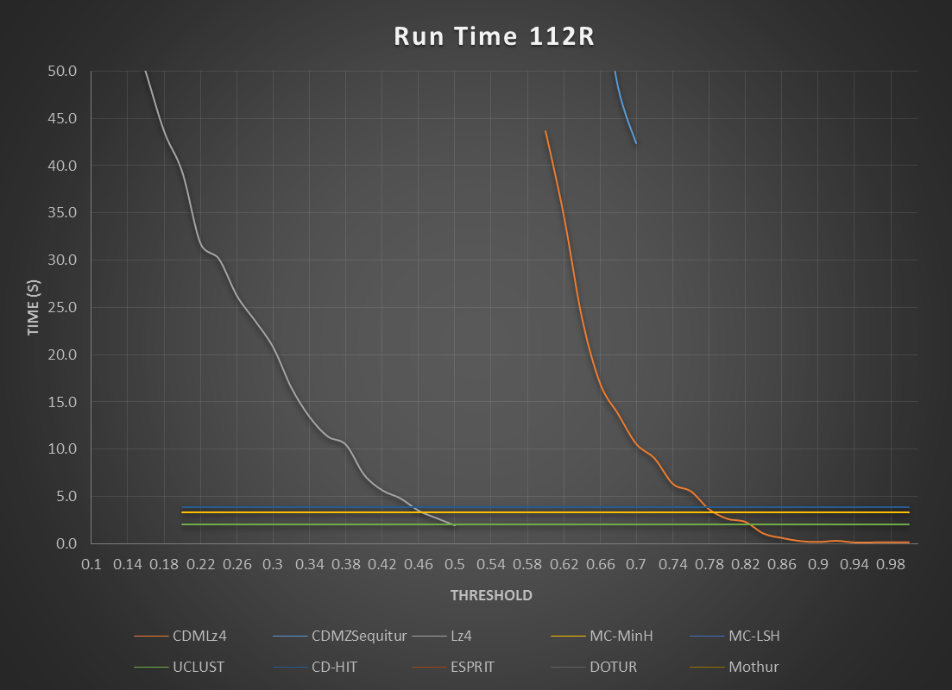
3% Error Data set

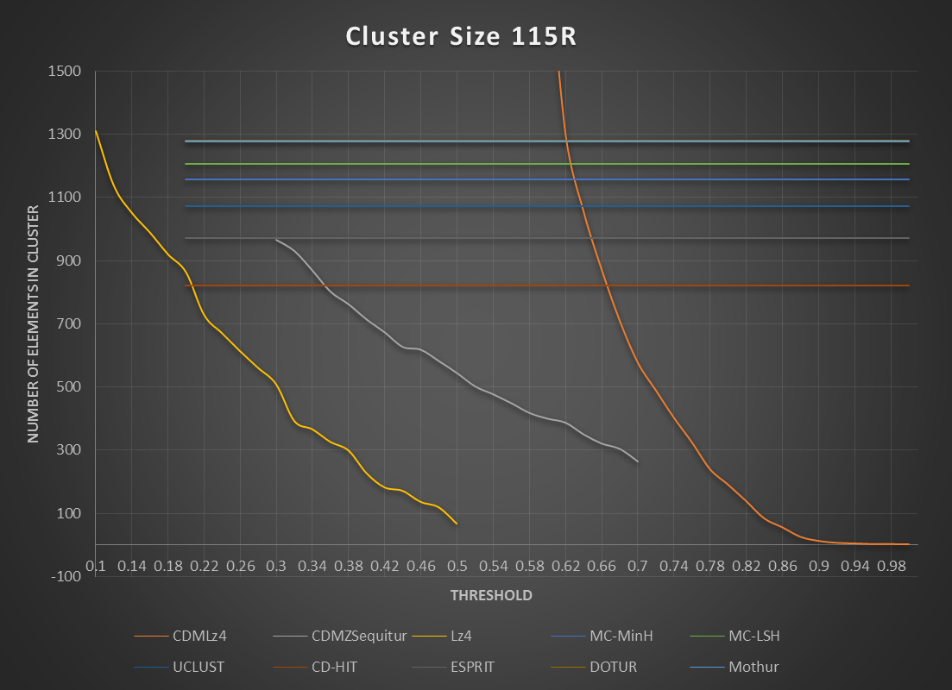
5% Error Data Set

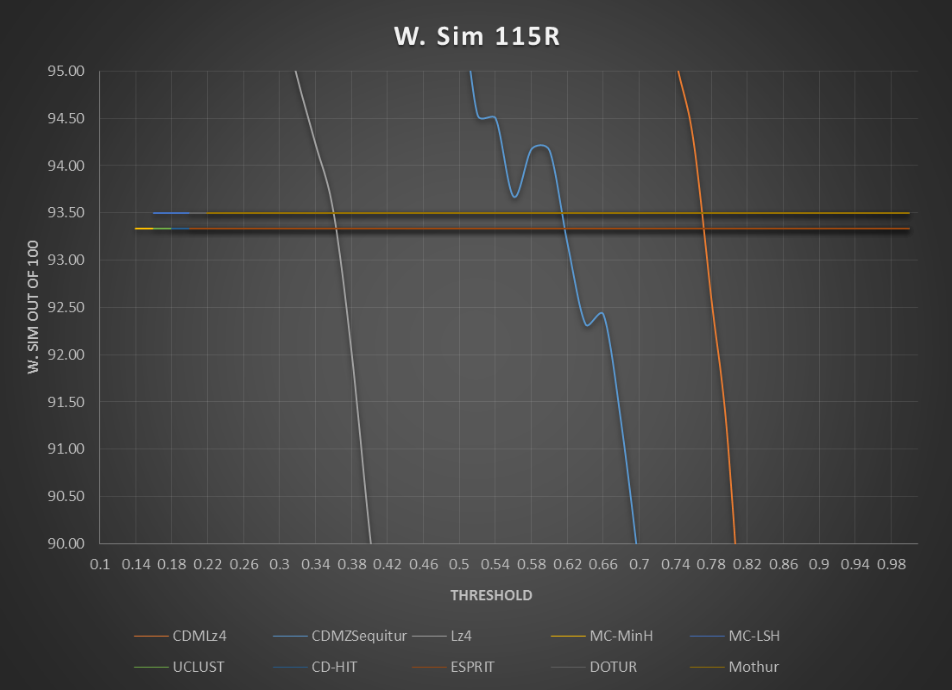
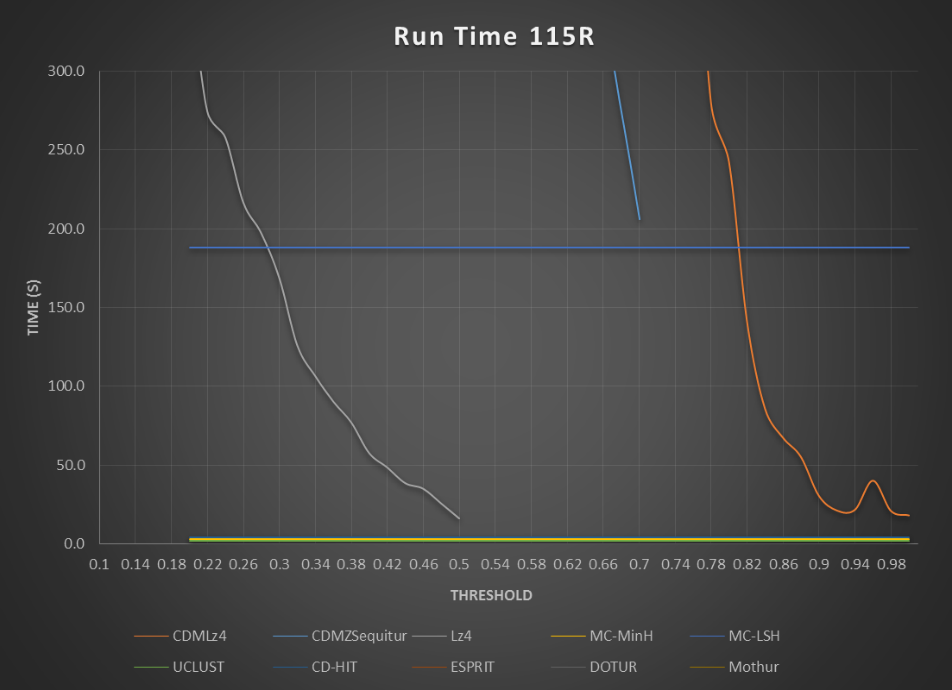
112R



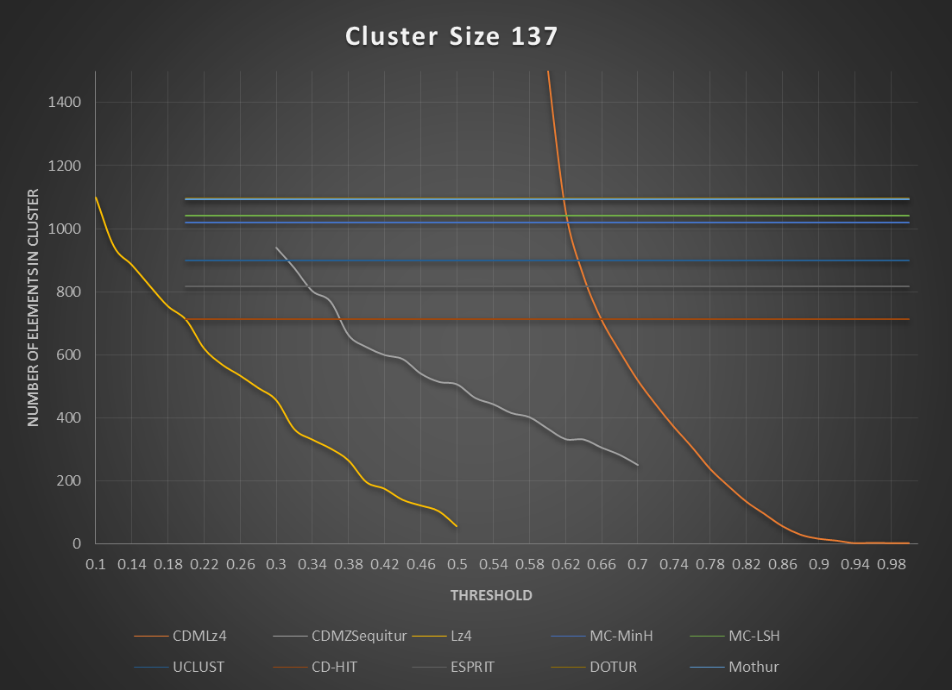
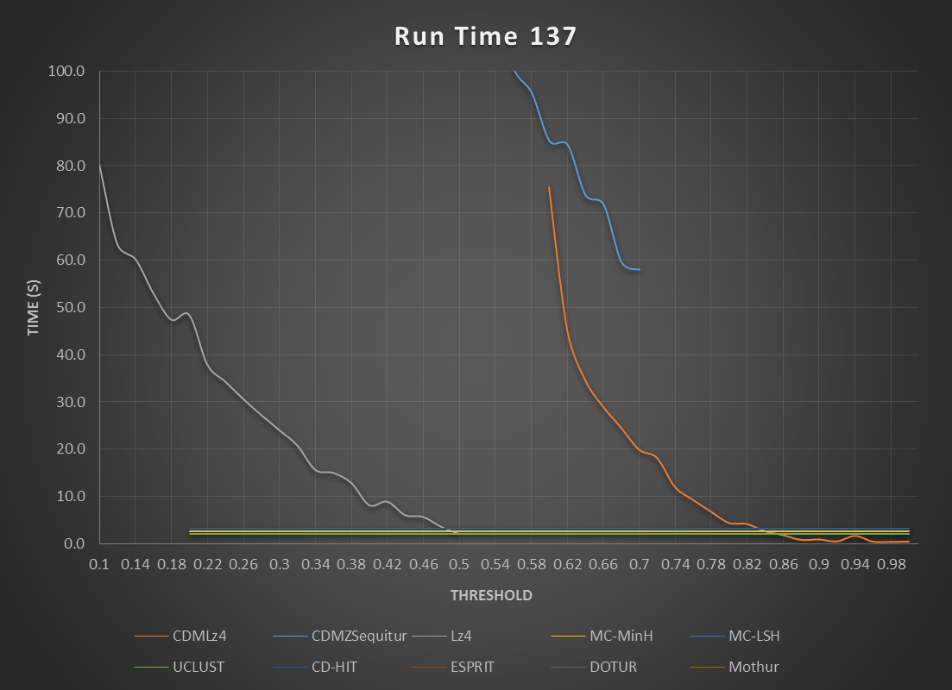


115R

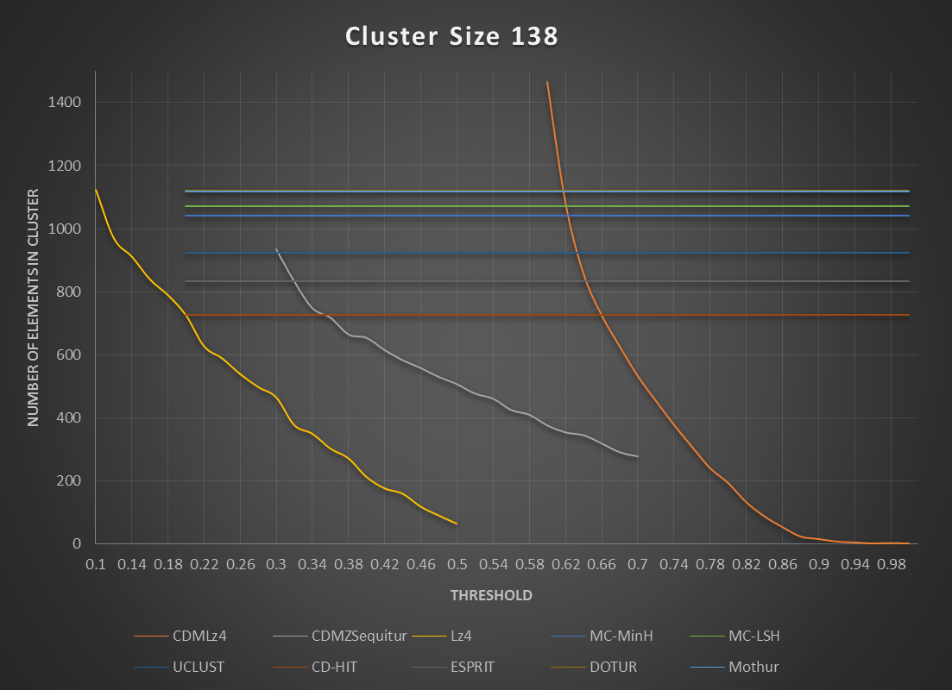




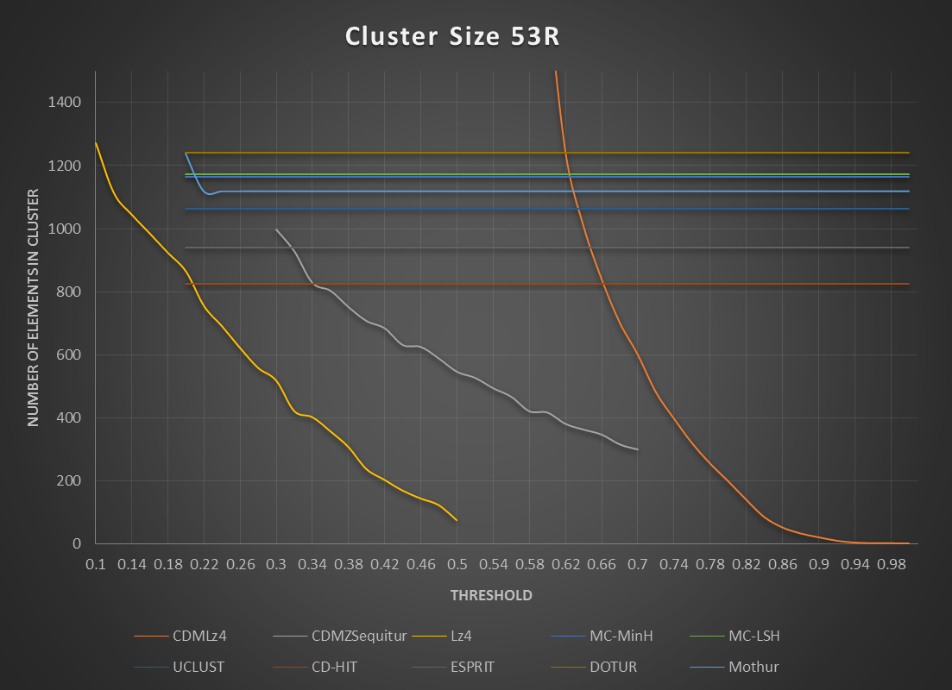
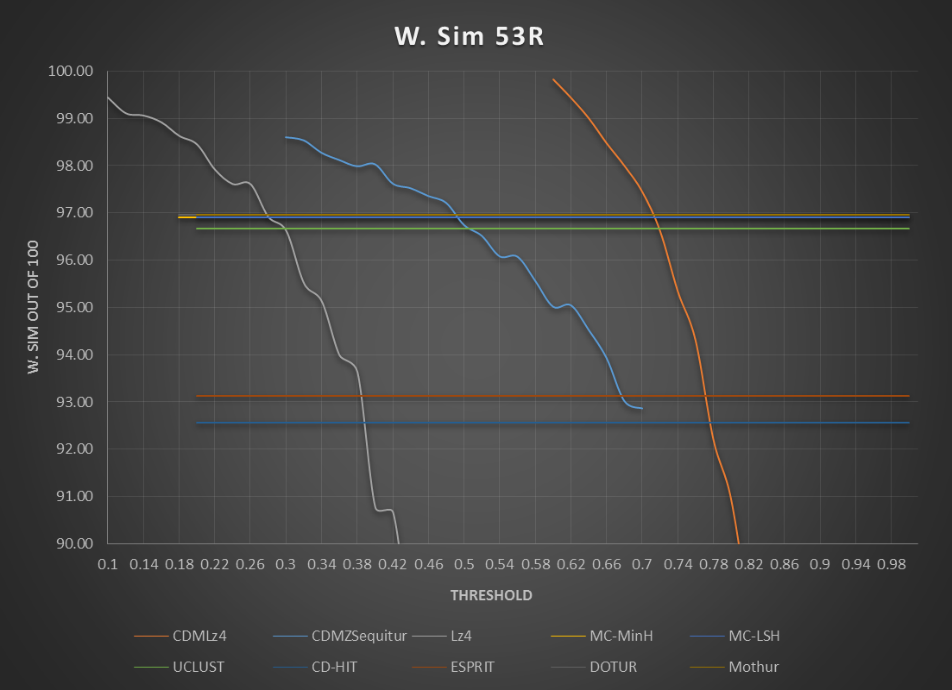
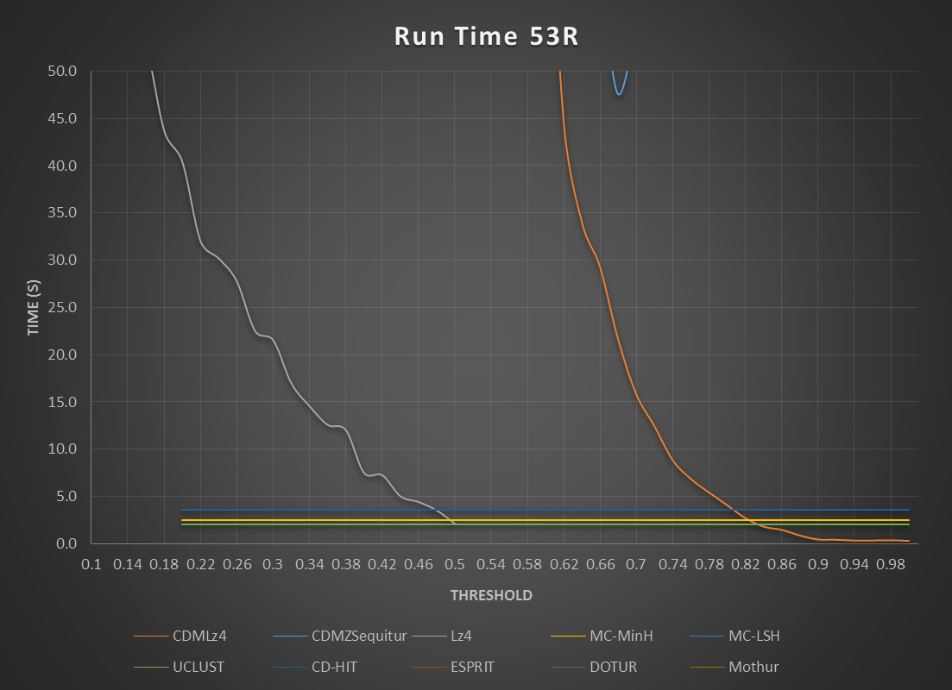
137



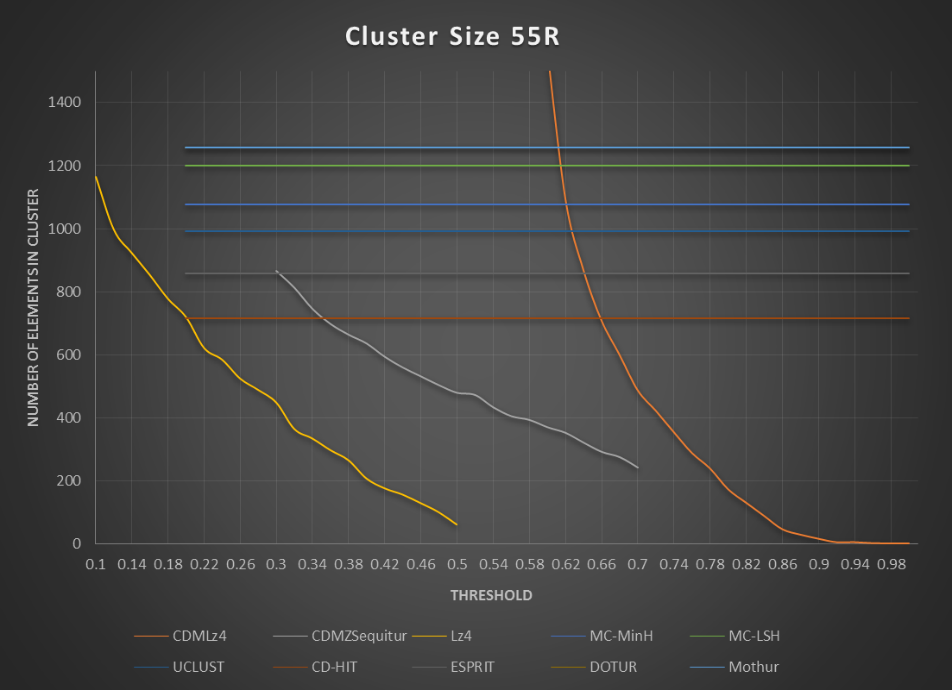
138



53R



55R



**Optimal Thresholds by W. Sim comparison**

|  |  |
| --- | --- |
| Algorithm | Threshold |
| CDMLz4 | .76 |
| CDMSequitur | .66 |
| Lz4 | .3 |

**Full Results**

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Algorithm | Metric | | 53R | 55R | 112R | 115R | 137 | 138 |
| CDMLz4 | # Clu | | 320 | 288 | 507 | 324 | 306 | 308 |
| W.Sim | | 94.30 | 93.35 | 92.01 | 94.30 | 93.81 | 93.91 |
| Time(s) | | 6.9 | 5.1 | 5.5 | 5.3 | 9.2 | 6.4 |
| CDMSequitur | # Clu | 346 | | 292 | 528 | 322 | 306 | 319 |
| W.Sim | 93.93 | | 93.07 | 92.41 | 92.42 | 93.79 | 93.99 |
| Time(s) | 59.4 | | 44.9 | 59.5 | 333.0 | 71.8 | 58.3 |
| Lz4 | # Clu | 517 | | 448 | 775 | 509 | 456 | 465 |
| W.Sim | 96.63 | | 95.39 | 96.00 | 95.73 | 96.08 | 95.91 |
| Time(s) | 21.5 | | 14.3 | 20.8 | 168.9 | 24.0 | 19.2 |
| MC-MinH | # Clu | 1165 | | 1077 | 1634 | 1156 | 1020 | 1042 |
| W.Sim | 96.90 | | 92.45 | 91.18 | 93.33 | 95.86 | 93.10 |
| Time(s) | 2.5 | | 2.1 | 3.3 | 3.0 | 2.7 | 2.5 |
| MC-LSH | #Clu | 1172 | | 1199 | 1795 | 1205 | 1041 | 1072 |
| W.Sim | 96.90 | | 93.12 | 91.33 | 93.50 | 95.86 | 93.10 |
| Time(s) | 161.0 | | 183.0 | 317.0 | 188.0 | 172.0 | 175.0 |
| UCLUST | #Clu | 1062 | | 992 | 1561 | 1071 | 900 | 923 |
| W.Sim | 96.67 | | 91.67 | 91.02 | 93.33 | 93.50 | 92.82 |
| Time(s) | 2.0 | | 2.0 | 2.0 | 2.0 | 2.0 | 2.0 |
| CD-HIT | #Clu | 824 | | 716 | 1196 | 820 | 712 | 725 |
| W.Sim | 92.56 | | 90.80 | 90.61 | 93.33 | 91.82 | 90.16 |
| Time(s) | 3.6 | | 3.1 | 3.9 | 3.8 | 3.2 | 3.1 |
| ESPRIT | #Clu | 940 | | 859 | 1361 | 970 | 818 | 832 |
| W.Sim | 93.12 | | 91.35 | 90.88 | 93.33 | 91.82 | 90.16 |
| Time(s) | 283.0 | | 266.0 | 537.0 | 348.0 | 280.0 | 296.0 |
| DOTUR | #Clu | 1241 | | 1258 | 1854 | 1279 | 1096 | 1121 |
| W.Sim | 96.95 | | 94.06 | 91.33 | 93.50 | 95.86 | 93.10 |
| Time(s) | 5129.0 | | 3511.0 | 5567.0 | 9237.0 | 6563.0 | 5618.0 |
| Mothur | #Clu | 1238 | | 1256 | 1853 | 1278 | 1094 | 1119 |
| W.Sim | 96.95 | | 94.06 | 91.33 | 93.50 | 95.86 | 93.10 |
| Time(s) | 10130.0 | | 5940.0 | 12303.0 | 13501.0 | 12861.0 | 12310.0 |

**Species Diversity comparison with MC-MinH**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| SID | Algorithm | Chao1 Index | Shannon Index | ACE Index |
| 53R | CDMLz4 | 1556.9 | 3.2 | 737.7 |
| CDMSequitur | 573.2 | 3.2 | 811.3 |
| Lz4 | 868.8 | 3.5 | 1267.0 |
| MC-MInH | 2276.3 | 4.4 | 2243.7 |
| 55R | CDMLz4 | 447.0 | 3.2 | 678.9 |
| CDMSequitur | 402.5 | 3.1 | 663.2 |
| Lz4 | 834.1 | 3.5 | 1093.5 |
| MC-MInH | 2182.8 | 4.6 | 2214.1 |
| 112R | CDMLz4 | 883.1 | 3.9 | 1256.1 |
| CDMSequitur | 925.2 | 3.9 | 1312.5 |
| Lz4 | 1471.7 | 4.4 | 2051.5 |
| MC-MInH | 3931.3 | 5.3 | 4202.7 |
| 115R | CDMLz4 | 494.5 | 3.2 | 762.6 |
| CDMSequitur | 463.9 | 3.1 | 749.5 |
| Lz4 | 834.5 | 3.4 | 1244.5 |
| MC-MInH | 2411.4 | 4.6 | 2455.8 |
| 137 | CDMLz4 | 430.0 | 3.5 | 679.6 |
| CDMSequitur | 393.8 | 3.5 | 667.6 |
| Lz4 | 714.4 | 3.9 | 1049.0 |
| MC-MInH | 1992.2 | 4.8 | 1800.1 |
| 138 | CDMLz4 | 437.4 | 3.2 | 709.3 |
| CDMSequitur | 476.1 | 3.3 | 749.9 |
| Lz4 | 658.7 | 3.6 | 1075.2 |
| MC-MInH | 1713.8 | 4.4 | 1760.3 |

**Optimal Thresholds by # Clu comparison**

|  |  |
| --- | --- |
| Algorithm | Threshold |
| CDMLz4 | .64 |
| CDMSequitur | .34 |
| Lz4 | .18 |

**Full Results**

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Algorithm | Metric | | 53R | 55R | 112R | 115R | 137 | 138 |
| CDMLz4 | # Clu | | 1009 | 868 | 1419 | 1056 | 849 | 858 |
| W.Sim | | 99.00 | 98.69 | 98.72 | 99.03 | 99.21 | 98.90 |
| Time(s) | | 34 | 24.2 | 23.9 | 24.1 | 34.7 | 27.8 |
| CDMSequitur | # Clu | 827 | | 745 | 1236 | 870 | 803 | 748 |
| W.Sim | 98.27 | | 98.14 | 97.94 | 97.94 | 98.74 | 98.49 |
| Time(s) | 187.7 | | 154.6 | 201.9 | 1032.5 | 244.3 | 186.7 |
| Lz4 | # Clu | 922 | | 777 | 1302 | 921 | 754 | 789 |
| W.Sim | 98.64 | | 98.26 | 98.50 | 98.52 | 98.66 | 98.66 |
| Time(s) | 43.7 | | 31.9 | 43.6 | 384.4 | 47.4 | 37.6 |
| MC-MinH | # Clu | 1165 | | 1077 | 1634 | 1156 | 1020 | 1042 |
| W.Sim | 96.90 | | 92.45 | 91.18 | 93.33 | 95.86 | 93.10 |
| Time(s) | 2.5 | | 2.1 | 3.3 | 3.0 | 2.7 | 2.5 |
| MC-LSH | #Clu | 1172 | | 1199 | 1795 | 1205 | 1041 | 1072 |
| W.Sim | 96.90 | | 93.12 | 91.33 | 93.50 | 95.86 | 93.10 |
| Time(s) | 161.0 | | 183.0 | 317.0 | 188.0 | 172.0 | 175.0 |
| UCLUST | #Clu | 1062 | | 992 | 1561 | 1071 | 900 | 923 |
| W.Sim | 96.67 | | 91.67 | 91.02 | 93.33 | 93.50 | 92.82 |
| Time(s) | 2.0 | | 2.0 | 2.0 | 2.0 | 2.0 | 2.0 |
| CD-HIT | #Clu | 824 | | 716 | 1196 | 820 | 712 | 725 |
| W.Sim | 92.56 | | 90.80 | 90.61 | 93.33 | 91.82 | 90.16 |
| Time(s) | 3.6 | | 3.1 | 3.9 | 3.8 | 3.2 | 3.1 |
| ESPRIT | #Clu | 940 | | 859 | 1361 | 970 | 818 | 832 |
| W.Sim | 93.12 | | 91.35 | 90.88 | 93.33 | 91.82 | 90.16 |
| Time(s) | 283.0 | | 266.0 | 537.0 | 348.0 | 280.0 | 296.0 |
| DOTUR | #Clu | 1241 | | 1258 | 1854 | 1279 | 1096 | 1121 |
| W.Sim | 96.95 | | 94.06 | 91.33 | 93.50 | 95.86 | 93.10 |
| Time(s) | 5129.0 | | 3511.0 | 5567.0 | 9237.0 | 6563.0 | 5618.0 |
| Mothur | #Clu | 1238 | | 1256 | 1853 | 1278 | 1094 | 1119 |
| W.Sim | 96.95 | | 94.06 | 91.33 | 93.50 | 95.86 | 93.10 |
| Time(s) | 10130.0 | | 5940.0 | 12303.0 | 13501.0 | 12861.0 | 12310.0 |

**Species Diversity comparison with MC-MinH**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| SID | Algorithm | Chao1 Index | Shannon Index | ACE Index |
| 53R | CDMLz4 | 1949.1 | 4.3 | 2140.1 |
| CDMSequitur | 1615.9 | 3.9 | 2147.8 |
| Lz4 | 1951.1 | 4.1 | 2424.1 |
| MC-MInH | 2276.3 | 4.4 | 2243.7 |
| 55R | CDMLz4 | 1766.1 | 4.4 | 2304.4 |
| CDMSequitur | 1499.3 | 4.1 | 1945.7 |
| Lz4 | 1619.6 | 4.1 | 2096.7 |
| MC-MInH | 2182.8 | 4.6 | 2214.1 |
| 112R | CDMLz4 | 3304.4 | 5.1 | 4479.5 |
| CDMSequitur | 2755.1 | 4.9 | 3942.1 |
| Lz4 | 3060.4 | 5.1 | 4296.9 |
| MC-MInH | 3931.3 | 5.3 | 4202.7 |
| 115R | CDMLz4 | 2087.3 | 4.3 | 2822.3 |
| CDMSequitur | 1908.4 | 4.0 | 2356.8 |
| Lz4 | 1999.9 | 4.1 | 2449.6 |
| MC-MInH | 2411.4 | 4.6 | 2455.8 |
| 137 | CDMLz4 | 1583.0 | 4.6 | 2054.3 |
| CDMSequitur | 1763.0 | 4.5 | 2024.1 |
| Lz4 | 1344.3 | 4.4 | 1813.2 |
| MC-MInH | 1992.2 | 4.8 | 1800.1 |
| 138 | CDMLz4 | 1557.2 | 4.2 | 2159.1 |
| CDMSequitur | 1354.4 | 4.0 | 1888.2 |
| Lz4 | 1305.1 | 4.1 | 1942.8 |
| MC-MInH | 1713.8 | 4.4 | 1760.3 |