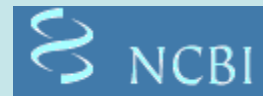
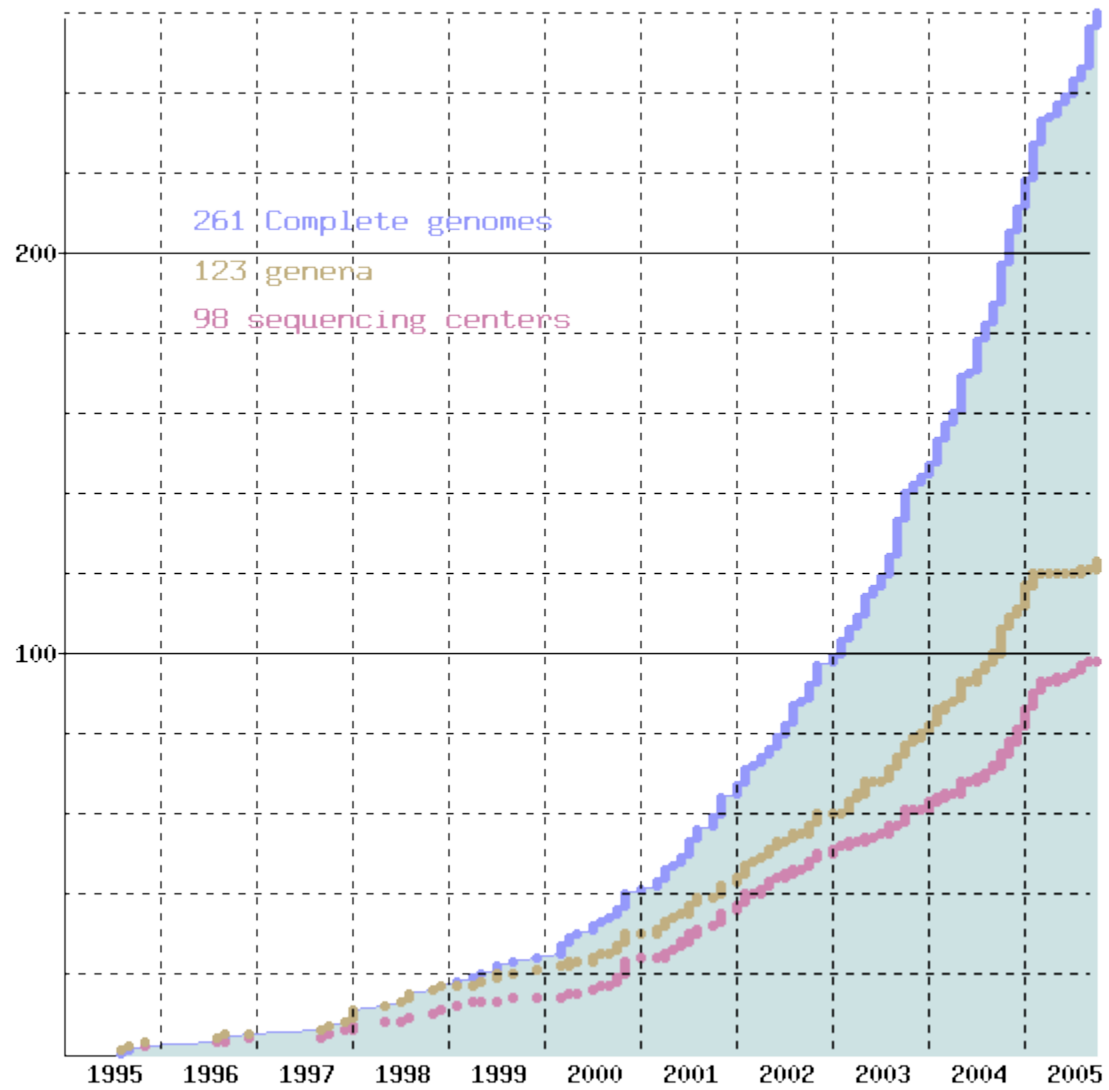


Tight clusters of proteins

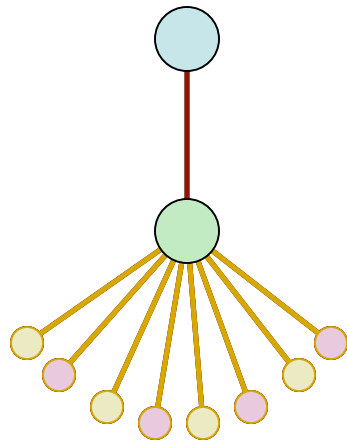
Boris Kiryutin
Roman Tatusov



The progress of complete genome sequencing



BLAST Uppermost Details (BUDs)



Inspect BLAST hits and find groups of top list proteins

BLAST hits

| sll0173 hits | |
|--------------|---|
| sll0173 | 1 |
| RSp0135 | 2 |
| PA0468 | 3 |
| RSc3155 | 4 |
| sll1049_1 | 5 |
| AGc1268 | 6 |
| RSc2755 | 7 |

Comparison of BUDs and COGs

| |
|---|
| 66 genomes |
| 192,987 proteins |
| 4,873 COGs |
| 35,678 BUDs |
| 139,866 proteins in BUDs |
| 1901 BUDs identical to COGs |
| 3334 COGs have more than 80% of proteins in BUDs |

Examples of BUDs for taxonomic orders

| <u>Rhizobiales</u> patterns in <u>BUDs</u> | | | | | | |
|--|------|------|------|-------------|------|------|
| # of proteins in genomes | | | | # of BUDs | | |
| Atu | Sme | Bme | Mlo | | | |
| 3407 | 3891 | 2210 | 3610 | 4919 | | |
| 3697 | 4389 | 2269 | 4299 | | | |
| 1478 | 1498 | 1432 | 1503 | <u>1417</u> | 1341 | 1088 |
| 181 | 183 | 176 | - | <u>173</u> | 156 | 138 |
| 618 | 613 | - | 639 | <u>562</u> | 523 | 353 |
| 54 | - | 53 | 65 | <u>53</u> | 43 | 37 |
| - | 161 | 151 | 161 | <u>150</u> | 134 | 119 |
| 791 | 792 | - | - | <u>729</u> | 629 | 387 |
| 118 | - | 118 | - | <u>116</u> | 100 | 95 |
| - | 70 | 69 | - | <u>68</u> | 62 | 58 |
| 304 | - | - | 315 | <u>288</u> | 254 | 191 |
| - | 746 | - | 760 | <u>664</u> | 579 | 371 |
| - | - | 208 | 224 | <u>207</u> | 179 | 159 |
| 153 | - | - | - | <u>69</u> | 51 | 46 |
| - | 326 | - | - | <u>128</u> | 99 | 79 |
| - | - | 62 | - | <u>26</u> | 22 | 18 |
| - | - | - | 632 | <u>269</u> | 180 | 140 |

| <u>Campylobacterales</u> patterns in <u>BUDs</u> | | | | | |
|--|------|-----|------------|-----|-----|
| # of proteins in genomes | | | # of BUDs | | |
| Hpy | jHp | Cje | | | |
| 1379 | 1377 | 757 | 1411 | | |
| 1463 | 1452 | 804 | | | |
| 730 | 731 | 737 | <u>725</u> | 658 | 607 |
| 715 | 714 | - | <u>647</u> | 411 | 341 |
| 2 | - | 2 | <u>2</u> | 2 | 2 |
| - | 3 | 4 | <u>3</u> | 2 | 2 |
| 16 | - | - | <u>5</u> | 3 | 2 |
| - | 4 | - | <u>2</u> | 0 | 0 |
| - | - | 61 | <u>27</u> | 17 | 12 |

| <u>Mycoplasmatales</u> patterns in <u>BUDs</u> | | | | | | |
|--|-----|-----|-----|------------|-----|-----|
| # of proteins in genomes | | | | # of BUDs | | |
| Uur | Mpu | Mpn | Mge | | | |
| 197 | 144 | 518 | 484 | 616 | | |
| 234 | 206 | 585 | 492 | | | |
| 18 | 18 | 18 | 18 | <u>17</u> | 15 | 15 |
| 1 | 1 | 1 | - | <u>1</u> | 1 | 1 |
| 123 | - | 120 | 120 | <u>119</u> | 109 | 106 |
| - | 34 | 34 | 33 | <u>33</u> | 32 | 31 |
| 45 | 48 | - | - | <u>40</u> | 27 | 27 |
| 2 | - | 2 | - | <u>2</u> | 2 | 2 |
| - | 13 | 14 | - | <u>13</u> | 11 | 11 |
| - | - | 342 | 321 | <u>315</u> | 237 | 221 |
| 45 | - | - | - | <u>18</u> | 7 | 6 |
| - | 92 | - | - | <u>40</u> | 17 | 15 |
| - | - | 54 | - | <u>18</u> | 3 | 2 |