

Ref: Ms. No. AMAB-D-11-00456

**Comparative Structural Bioinformatics Analysis of
Bacillus amyloliquefaciens Chemotaxis Proteins within
Bacillus subtilis Group**

Anna Yssel¹, Oleg Reva² and Özlem Tastan Bishop¹

¹Rhodes University Bioinformatics (RUBi), Department of Biochemistry, Microbiology and Biotechnology, Rhodes University, Grahamstown, 6140, South Africa

²Bioinformatics and Computational Biology Unit, Department of Biochemistry, University of Pretoria, Pretoria, 0002, South Africa

Table S1: Selecton optimized parameters, Bayesian Ka/Ks ratio results and results of comparison between MEC and M8a models.

| Protein | Model | Selecton optimized parameters | Outcome of model: Positively selected sites * | AIC score** |
|-------------|-------|---|--|----------------|
| CheA | MEC | log-likelihood: -8271.08 (tv): 4.38708 (tr): 3.43969 f=prob(selection): 1 α : 0.211348 β : 2.65081 | No positive selection detected. | Not applicable |
| CheB | MEC | log-likelihood: -5109.71 (tv): 4.53682 (tr): 3.37548 f=prob(selection): 1 α : 0.328257 β : 2.72201 | 25, 90, 277, 279 | 10229.47 |
| CheB | M8a | log-likelihood: -5214.74 κ : 1.62 α : 0.560397 β : 2.70564 ω s: 1 prob(additional omega category): 0.0172209 | Significance test passed | 10437.51 |
| CheC | MEC | log-likelihood: -2456.22 (tr): 4.39392 (tv): 3.52879 f=prob(selection): 1 α : 0.260817 β : 2.65081 | 86 | 4922.53 |
| CheC | M8a | log-likelihood: -2484.41 κ : 1.35981 α : 0.360544 β : 2.70564 ω s: 1 prob(additional omega category): 0.0172209 | Significance test passed | 4976.88 |
| CheD | MEC | log-likelihood: -2003.96 (tv): 4.32103 (tr): 3.49261 f=prob(selection): 1 α : 0.230654 β : 2.65081 | No positive selection detected. | Not applicable |
| CheR | MEC | log-likelihood: -3499.48 (tr): 4.51696 (tv): 4.4594 f=prob(selection): 1 α : 0.248046 β : 2.65081 | No positive selection detected | Not applicable |

| Protein | Model | Selecton optimized parameters | Outcome of model: Positively selected sites * | AIC score** |
|-------------|-------|---|--|----------------|
| CheV | MEC | log-likelihood: -3712.24 (tr): 4.54706 (tv): 3.83795 f=prob(selection): 1 α : 0.238235 β : 2.65081 | 26 | 7434.54 |
| CheV | M8a | log-likelihood: -3772.01 κ : 1.31927 α : 0.371049 β : 2.70564 ω s: 1 prob(additional omega category): 0.0172209 | Significance test passed | 7552.06 |
| CheW | MEC | log-likelihood: -2043.23 (tv): 4.50976 (tr): 3.37874 f=prob(selection): 1 α : 0.323541 β : 2.72201 | 18 | 4096.58 |
| CheW | M8a | log-likelihood: -2080.8 κ : 1.458 α : 0.540009 β : 2.70564 ω s: 1 prob(additional omega category): 0.0172209 | Significance test passed | 4169.68 |
| CheY | MEC | log-likelihood: -1145.55 (tv): 4.19899 (tr): 2.52786 f=prob(selection): 1 α : 0.122291 β : 2.65081 | No positive selection detected | Not applicable |
| McpA | MEC | log-likelihood: -7534.84 (tr): 4.53548 (tv): 3.04065 f=prob(selection): 1 α : 0.302022 β : 2.72201 | 95, 99, 112, 234, 235, 244, 289 | 15079.71 |
| McpA | M8a | log-likelihood: -7665.58 κ : 1.62 α : 0.444086 β : 2.70564 ω s: 1 prob(additional omega category): 0.0172209 | Significance test passed | 15339.18 |
| McpB | MEC | log-likelihood: -9747.81 (tr): 4.57997 (tv): 2.87187 | 2, 92, 99, 288, 376, 408, 412, 478 | 19505.65 |

| Protein | Model | Selecton optimized parameters | Outcome of model: Positively selected sites * | AIC score** |
|-------------|-------|---|--|----------------|
| McpB | M8a | f=prob(selection): 0.854102 α: 0.324162 β: 2.69105 log-likelihood: -9828.76 κ: 1.62 α: 0.653015 β: 2.70564 ωs: 1 prob(additional omega category): 0.0181427 | Significance test passed | 19665.54 |
| McpC | MEC | log-likelihood: -9903.82 (tv): 4.21725 (tr): 3.28392 f=prob(selection): 1 α: 0.408259 β: 2.68488 | 52, 71, 123, 169, 171, 242, 276, 322, 354, 536, 647, 651 | 19817.67 |
| McpC | M8a | log-likelihood: -10060 κ: 1.62 α: 0.633274 β: 2.57959 ωs: 1 prob(additional omega category): 0.0172209 | Significance test passed | 20128.02 |
| gyrA | MEC | log-likelihood: -8975.74 (tv): 4.3907 (tr): 2.95171 f=prob(selection): 1 α: 0.122291 β: 2.65081 | No positive selection detected | Not applicable |

* Amino acid sites are numbered according to their position in the reference sequence from *B. amyloliquefaciens* FZB 42, The M8a null model was not run when no positive selection was detected by the MEC model.

** Statistical significance test passed if AIC score of MEC model < AIC score M8a model.

Parameters of the M8a model are: α and β are the shape parameters of the β distribution. K is the transition/transversion ratio. Ω_s is the additional category representing positive selection (ω_s set to 1). p₁ is the proportion of ω_s. Parameters of MEC model are: α, β, tr (rate of transition), tv (rate of transversion), f (proportion of sites under no selection).

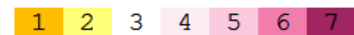
Table S2: K_a/K_s scores for sites detected to be under positive selection as obtained under the MEC and M8a models. A normalized value for $K_a/K_s > 1$ indicates positive selection.

| Protein | Residue | MEC: K_a^1/K_s^1 | M8a: K_a^0/K_s^0 | Normalized= $(K_a^1/K_s^1)/$ (K_a^0/K_s^0) |
|---------|---------|--------------------|--------------------|--|
| CheB | 25, | 1.1 | 0.58 | 1.90 |
| CheB | 90, | 1.2 | 0.57 | 2.11 |
| CheB | 277, | 1.2 | 0.68 | 1.76 |
| CheB | 279 | 1.1 | 0.59 | 1.86 |
| CheC | 86 | 1.3 | 0.8 | 1.63 |
| CheV | 26 | 1.1 | 0.71 | 1.55 |
| CheW | 18 | 1.2 | 0.61 | 1.97 |
| McpA | 95 | 1.4 | 0.85 | 1.65 |
| McpA | 99 | 1.2 | 0.76 | 1.58 |
| McpA | 112 | 1.1 | 0.47 | 2.34 |
| McpA | 234 | 1.3 | 0.8 | 1.6 |
| McpA | 235 | 1.2 | 0.72 | 1.67 |
| McpA | 244 | 1.1 | 0.45 | 2.44 |
| McpA | 289 | 1.1 | 0.59 | 1.86 |
| McpB | 2 | 1.3 | 0.64 | 2.03 |
| McpB | 92 | 1.3 | 0.67 | 1.94 |
| McpB | 99 | 1.1 | 0.52 | 2.12 |
| McpB | 288 | 1.1 | 0.68 | 1.62 |
| McpB | 376 | 1.1 | 0.5 | 2.2 |
| McpB | 408 | 1.1 | 0.67 | 1.64 |
| McpB | 412 | 1.1 | 0.56 | 1.96 |
| McpB | 478 | 1.3 | 0.84 | 1.55 |
| McpC | 52 | 1.2 | 0.47 | 2.55 |
| McpC | 71 | 1.4 | 0.65 | 2.15 |
| McpC | 123 | 1.1 | 0.61 | 1.8 |
| McpC | 169 | 1.4 | 0.59 | 2.37 |
| McpC | 171 | 1.1 | 0.5 | 2.2 |
| McpC | 242 | 1.5 | 0.69 | 2.17 |
| McpC | 276 | 1.1 | 0.51 | 2.16 |
| McpC | 322 | 1.1 | 0.57 | 1.93 |
| McpC | 354 | 1.4 | 0.74 | 1.89 |
| McpC | 536 | 1.3 | 0.67 | 1.94 |
| McpC | 647 | 1.2 | 0.63 | 1.9 |
| McpC | 651 | 1.2 | 0.69 | 1.74 |

1 11 21 31 41
 MSEQNTPQVR EVNISQEMRT SFLDYAMSVI VSRALPDVRD GLKPVHRRIL
 51 61 71 81 91
 YAMNDLGMTS DKPYKKSARI VGEVIGKYHP HGDSAVYESM VRMAQDFNYR
 101 111 121 131 141
 YMLVDGHGNF GSVVDGDSAAA MRYTEARMSK IAMEILRDIT KDTIDYQDNY
 151 161 171 181 191
 DGAEREPVVM PSRFPNLLVN GAAGIAVGMA TNIPPHQLGE VIEGVLAVSE
 201 211 221 231 241
 NPEITNQELM EYIPGPDFPT AGQILGRSGI RKAYESGRGS ITIRAKAEIE
 251 261 271 281 291
 ETS SGKERII VTELPYQVNK ARLIEKIADL VRDKKIEGIT DLRDESDRNG
 301 311 321 331 341
 MRIVIEIRRD ANAHVILNNL YKQTALQTSF GINLLALVDG QPKVLSLKQC
 351 361 371 381 391
 LEHYLDHQKV VIRRRRTAYEL RKAEARAHIL EGLRIALDHL DAVISLIRNS
 401 411 421 431 441
 QTAEIARTGL IEQFSLTEKQ AQAILDMRLQ RL TGLEREKI EEEYQSLVAL
 451 461 471 481 491
 I AELKDI LAN EARVLEI IRE ELNEIKERFN DERRTEIVTS GLETIEDEDL
 501 511 521 531 541
 IERENIVITL THNGYVKRLP ASTYRSQKRG GKGVQGMGTN EDDFVEHLIS
 551 561 571 581 591
 TSTHDTILFF SNKKGKRYRSK GYEIPEYGR T AKGIPTIINLL EVEKGEWINA
 601 611 621 631 641
 IIPVSTFDEE QYLFFTTKQG VSKRTALSQF ANIRNNGLIA LGLREDDLM
 651 661 671 681 691
 AVRITDGKKQ I IIGTKNGLL IRFPEEDVRQ MGR TAAGVKG ITLTDDDVVV
 701 711 721 731 741
 GMEILEEDSH VLI V TENG YG KRT PASEYRV QSRGGKGLKT CKITDSNGPL
 751 761 771 781 791
 VTVKATK GEE DLM IITASGV LIRMDINDIS TTGRVTQGVRL IRMSDQEHV
 801 811
 ATVALVEKNE EEPEETEEV

Legend:

The selection scale:



Positive selection Purifying selection

Fig. S1: Selecton results under the MEC model, colour coded onto the GyrA amino acid sequence of *B. amyloliquefaciens* FZB42.