Common gene arrangements

|  |  |  |
| --- | --- | --- |
| Genes present | Strains | Contigs/indeterminate? |
| PknH, tbD2, pknH2, embR | Remaining 20 strains | 1, 1, 4, 3, 5, 4, 2, 17, 3, 6, 3, 3, 2, 2, 4, 2, 4, 4, 2, 2, |
| PknH, embR, BQ2027\_MB1299C, BQ2027\_MB1300C, pknH2, tbD2 | K85 (000665865) | Scaffold = 6 |
| PknH, embR, BQ2027\_MB1299C, BQ2027\_MB1300C, tbD2, pknH2 separate to pknH | MAL020193 | Scaffold = 7 |
| PknH, BQ2027\_MB1296, BQ2027\_MB1295, amiB2 | MAL010074 | Scaffold = 4 |
| PknH2, tbD2, pknH, BQ2027\_MB1296, BQ2027\_MB1295, amiB2 | MAL010081, MAL010131, MAL020176 | 081 = 2  131 = 3  176 = 2 |
| PknH, tbD2, pknH2, pknH, embR, BQ2027\_MB1299C, BQ2027\_MB1300C | MAL010118, MAL010129, 25 (third pknH gene, not at same location) | 118 = 5  129 = 3  25 = 1 |
| PknH, embR, BQ2027\_MB1299C, BQ2027\_MB1300C (no pknH2!) | UT307 | Scaffold = 1  No pknH2? |
| PknH, BQ2027\_MB1296, BQ2027\_MB1295,  tbD2, pknH2 (separate) | ATCC 25420 | Scaffold = 150  Small contig with tbD2 and pknH2 |
| AmiB2, BQ2027\_MB1295, BQ2027\_MB2096, pknH  PknH2, tbD2 (separate) | NLA000017458(SAMN14601854),  NLA000017458(SAMN15217203) | Scaffold = 90  Small contig with pknH2 and tbD2  Scaffold = 97 |
| AmiB2, BQ2027\_MB1295, BQ2027\_MB1296, pknH  TbD2, pknH2 (separate) | L12222(SAMN15217203) | Scaffold =142  Small contig with tbD2 and pknH2 |
| PknH, BQ2027\_MB11296, BQ2027\_MB1295, amiB2  PknH2, tbD2 (separate) | L12222(SAMN15217201) | Scaffold = 160  Small contig with pknH2 and tbD2 |

Protein sequence alignments – orthologues function – where do residues differ

|  |  |  |
| --- | --- | --- |
| Standard pknH protein coding sequence | GCA\_000159775.1\_ASM15977v1\_genomic  >EADCCBD\_01835 PKNH\_MYCBO Serine/threonine-protein kinase PknH  MSDAQDSRVGSMFGPYHLKRLLGRGGMGEVYEAEHTVKEWTVAVKLMTAEFSKDPVFRER  MKREARIAGRLQEPHVVPIHDYGEVDGQMFLEMRLVEGTDLDSVLKRFGPLTPPRAVAII  TQIASALDAAHADGVMHRDVKPQNILITRDDFAYLVDFGIASATTDEKLTQLGTAVGTWK  YMAPERFSNDEVTYRADIYALACVLHECLTGAPPYRADSAGTLVSSHLMGPIPQPSAIRP  GIPKAFDAVVARGMAKKPEDRYASAGDLALAAHEALSDPDQDHAADILRRSQESTLPAPP  KPVPPPTMPATAMAPRQPPAPPVIPPGVQPAPKPSYTPPAQPGPAGQRPGPTGQPSWAPN  SGPMPASGPTPTPQYYQGGGWGAPPSGGPSPWAQTPRKTNPWPLVAGAAAVVLVLVLGAI  GIWIAIRPKPVQPPQPVAEERLSALLLNSSEVNAVMGSSSMQPGKPITSMDSSPVTVSLP  DCQGALYTSQDPVYAGTGYTAINGLISSEPGDNYEHWVNQAVVAFPTADKARAFVQTSAD  KWKNCAGKTVTVTNKAKTYRWTFADVKGSPPTITVIDTQEGAEGWECQRAMSVANNVVVD  VNACGYQITNQAGQIAAKIVDKVNKE | K85, MAL010074 (and no pknH2),  MAL010081,  MAL010118 AND MAL010129(one of the two pknH genes has this sequence),  MAL010131,  MAL020176,  MAL020193– same pknH protein coding sequence (diff. pknH2), UT307(same pknH, no pknH2) |
| Standard pknH2 protein coding sequence | GCA\_000159775.1\_ASM15977v1\_genomic  >EADCCBD\_01837 A0A1R3XXU5\_MYCBO PknH2 protein  MSDAQDSRVGSMFGPYHLKRLLGRGGMGEVYEAEHTVKEWTVAVKLMTAEFSKDPVFRER  MKREARIAGRLQEPHVVPIHDYGEVDGQMFLEMRLVEGTDLDSVLKRFGPLTPPRAVAII  TQIASALDAAHADGVMHRDVKPQNILITRDDFAYLVDFGIASATTDEKLTQLGTAVGTWK  YMAPERFSNDEVTYRADSAGTLVSSHLMGPIPQPSAIRPGIPKAFDAVVARGMAKKPEDR  YASAGDLALAAHEALSDPDQDHAADILRRSQESTLPGTAAVTAQPPTMPTVTPPPIQAAP  TGQPSWAPNSGPMPASGPTPTPQYYQGGGWGAPPSGGPSPWAQTPRKTNPWPFVAVAAAV  VLVLVLGAIGIWIANRPDDNPKRNIATSPGTPTTTATTSLPATTTPTTAPASDPQTRLLS  MLPSGYPTGTCKPTTPKPNSIWVNAVAMVDCGQNTNQGGPSRAIYGLFANPDKLKQAFND  DIAAVELMNCPGEGPSPDGWHYNQTPDVTAGMIACGTYKNRPNVIWSNEAKLTLSDVFGD  PATIEDLHNWWAKYG |  |
| TbD2 |  |  |
| K85 pknH2 | GCA\_000665865.1\_Myco\_afri\_K85\_V1\_genomic  >FADCDBD\_04098 A0A1R3XXU5\_MYCBO PknH2 protein  MAVAAAVVLVLVLGAIGIWIANRPDDNPKRNIATSPGTPTTTATTSLPATTTPTTAPASD  PQTRLLSMLPSGYPTGTCKPTTPKPNSIWVNAVAMVDCGQNTNQGGPSRAIYGLFANPDK  LKQAFNDDIAAVELMNCPGEGPSPDGWHYNQTPDVTAGMIACGTYKNRPNVIWSNEAKLT  LSDVFGDPATIEDLHNWWAKYG |  |
| MAL010118  pknH | GCA\_000666625.1\_Myco\_afri\_MAL010118\_V1\_genomic  >BEDDBAC\_01357 PKNH\_MYCBO Serine/threonine-protein kinase PknH  MSDAQDSRVGSMFGPYHLKQLLGRGGMGEVYEAEHTVKEWTVAVKLMTAEFSKDPVFRER  MKREARIAGRLQEPHVVPIHDYGEVDGQMFLEMRLVEGTDLDSVLKRFGPLTPPRAVAII  TQIASALDAAHADGVMHRDVKPQNILITRDDFAYLVDFGIASATTDEKLTQLGTAVGTWK  YMAPERFSNDEVTYRADIYALACVLHECLTGAPPYRADSAGTLVSSHLMGPIPQPSAIRP  GIPKAFDAVVARGMAKKPEDRYASAGDLALAAHEALSDPDQDHAADILRRSQESTLPAPR  R | MAL010129 has the same protein coding sequence for its pknH |
| MAL010118  PknH2 | GCA\_000666625.1\_Myco\_afri\_MAL010118\_V1\_genomic  >BEDDBAC\_01356 A0A1R3XXU5\_MYCBO PknH2 protein  MPTVTPPPIQAAPTGQPSWAPNSGPMPASGPTPTPQYYQGGGWGAPPSGGPSPWAQTPRK  TNPWPFVAVAAAVVLVLVLGAIGIWIANRPDDNPKRNIATSPGTPTTTATTSLPATTTPT  TAPASDPQTRLLSMLPSGYPTGTCKPTTPKPNSIWVNAVAMVDCGQNTNQGGPSRAIYGL  FANPDKLKQAFNDDIAAVELMNCPGEGPSPDGWHYNQTPDVTAGMIACGTYKNRPNVIWS  NEAKLTLSDVFGDPATIEDLHNWWAKYG | 25 has the same protein coding sequence for its pknH2  MAL010129 has the same protein coding sequence for its pknH2 |
| MAL020193  PknH2 | GCA\_000667525.1\_Myco\_tube\_MAL020193\_V1\_genomic  >CBAADDB\_04092 A0A1R3XXU5\_MYCBO PknH2 protein  MGPIPQPSAIRPGIPKAFDAVVARGMAKKPEDRYASAGDLALAAHEALSDPDQDHAADIL  RRSQESTLPGTAAVTAQPPTMPTVTPPPIQAAPTGQPSWAPNSGPMPASGPTPTPQYYQG  GGWGAPPSGGPSPWAQTPRKTNPWPFVAVAAAVVLVLVLGAIGIWIANRPDDNPKRNIAT  SPGTPTTTATTSLPATTTPTTAPASDPQTRLLSMLPSGYPTGTCKPTTPKPNSIWVNAVA  MVDCGQNTNQGGPSRAIYGLFANPDKLKQAFNDDIAAVELMNCPGEGPSPDGWHYNQTPD  VTAGMIACGTYKNRPNVIWSNEAKLTLSDVFGDPATIEDLHNWWAKYG | ATCC25420  PknH2 has the same protein coding sequence |
| 25  PknH (three pknHs) | GCA\_001544855.1\_ASM154485v1\_genomic  >CDBCAFF\_01577 PKNH\_MYCBO Serine/threonine-protein kinase PknH  MSDAQDSRVGSMFGPYHLKRLLGRGGMGEVYEAEHTVKEWTVAVKLMTAEFSKDPVFRER  MKREARIAGRLQEPHVVPIHDYGEVDGQMFLEMRLVEGTDLDSVLKRFGPLTPPRAVAII  TQIASALDAAHADGVMHRDVKPQNILITRDDFAYLVDFGIASATTDEKLTQLGTAVGTWK  YMAPNGSPTTR  GCA\_001544855.1\_ASM154485v1\_genomic  >CDBCAFF\_01581 PKNH\_MYCBO Serine/threonine-protein kinase PknH  MSDAQDSRVGSMFGPYHLKRLLGRGGMGEVYEAEHTVKEWTVAVKLMTAEFSKDPVFRER  MKREARIAGRLQEPHVVPIHDYGEVDGQMFLEMRLVEGTDLDSVLKRFGPLTPPRAVAII  TQIASALDAAHADGVMHRDVKPQNILITRDDFAYLVDFGIASATTDEKLTQLGTAVGTWK  YMAPERFSNDEVTYRADSAGTLVSSHLMGPIPQPSAIRPASPKPLTRWSPAAWPKSPKTA  TPAPATWPWPPTKRSATPTKTTPPTSCAAAKNPPCPAPRR  GCA\_001544855.1\_ASM154485v1\_genomic  >CDBCAFF\_01575 PKNH\_MYCBO Serine/threonine-protein kinase PknH  MAGAAAVVLVLVLGAIGIWIAIRPKPVQPPQPVAEERLSALLLNSSEVNAVMGSSSMQPG  KPITSMDSSPVTVSLPDCQGALYTSQDPVYAGTGYTAINGLISSEPGDNYEHWVNQAVVA  FPTADKARAFVQTSADKWKNCAGKTVTVTNKAKTYRWTFADVKGSPPTITVIDTQEGAEG  WECQRAMSVANNVVVDVNACGYQITNQAGQIAAKIVDKVNKE |  |
| ATCC 25420  pknH | GCA\_002982335.1\_ASM298233v1\_genomic  >FBAFBCA\_01065 PKNH\_MYCBO Serine/threonine-protein kinase PknH  MLHECLTGAPPYRADSAGTLVSSHLMGPIPQPSAIRPGIPKAFDAVVARGMAKKPEDRYA  SAGDLALAAHEALSDPDQDHAADILRRSQESTLPAPPKPVPPPTMPATAMAPRQPPAPPV  IPPGVQPAPKPSYTPPAQPGPAGQRPGPTGQPSWAPNSGPMPASGPTPTPQYYQGGGWGA  PPSGGPSPWAQTPRKTNPWPLVAGAAAVVLVLVLGAIGIWIAIRPKPVQPPQPVAEERLS  ALLLNSSEVNAVMGSSSMQPGKPITSMDSSPVTVSLPDCQGALYTSQDPVYAGTGYTAIN  GLISSEPGDNYEHWVNQAVVAFPTADKARAFVQTSADKWKNCAGKTVTVTNKAKTYRWTF  ADVKGSPPTITVIDTQEGAEGWECQRAMSVANNVVVDVNACGYQITNQAGQIAAKIVDKV  NKE |  |
| NLA000017458(SAMN14601854)  pknH | GCA\_012584335.1\_ASM1258433v1\_genomic  >BDBDBAF\_01426 PKNH\_MYCBO Serine/threonine-protein kinase PknH  MAKKPEDRYASAGDLALAAHEALSDPDQDHAADILRRSQESTLPAPPKPVPPPTMPATAM  APRQPPAPPVTPPGVQPAPKPSYTPPAQPGPAGQRPGPTGQPSWAPNSGPMPASGPTPTP  QYYQGGGWGAPPSGGPSPWAQTPRKTNPWPLVAGAAAVVLVLVLGAIGIWIAIRPKPVQP  PQPVAEERLSALLLNSSEVNAVMGSSSMQPGKPITSMDSSPVTVSLPDCQGALYTSQDPV  YAGTGYTAINGLISSEPGDNYEHWVNQAVVAFPTADKARAFVQTSADKWKNCAGKTVTVT  NKAKTYRWTFADVKGSPPTITVIDTQEGAEGWECQRAMSVANNVVVDVNACGYQITNQAG  QIAAKIVDKVNKE | NLA00017458(SAMN15217203) has the same pknH protein coding sequence |
| NLA00017458(SAMN14601854)  pknH2 | GCA\_012584335.1\_ASM1258433v1\_genomic  >BDBDBAF\_04102 A0A1R3XXU5\_MYCBO PknH2 protein  MLVLVLGAIGIWIANRPDDNPKRNIATSPGTPTTTATTSLPATTTPTTAPASDPQTRLLS  MLPSGYPTGTCKPTTPKPNSIWVNAVAMVDCGQNTNQGGPSRAIYGLFANPDKLKQAFND  DIAAVELMNCPGEGPSPDGWHYNQTPDVTAGMIACGTYKNRPNVIWSNEAKLTLSDVFGD  PATIEDLHNWWAKYG |  |
| L12222 (SAMN15217203)  pknH | GCA\_013018365.1\_ASM1301836v1\_genomic  >ABBFEAE\_01417 PKNH\_MYCBO Serine/threonine-protein kinase PknH  MLVLVLGAIGIWIAIRPKPVQPPQPVAEERLSALLLNSSEVNAVMGSSSMQPGKPITSMD  SSPVTVSLPDCQGALYTSQDPVYAGTGYTAINGLISSEPGDNYEHWVNQAVVAFPTADKA  RAFVQTSADKWKNCAGKTVTVTNKAKTYRWTFADVKGSPPTITVIDTQEGAEGWECQRAM  SVANNVVVDVNACGYQITNQAGQIAAKIVDKVNKE |  |
| L12222  (SAMN15217203)  PknH2 | GCA\_013018365.1\_ASM1301836v1\_genomic  >ABBFEAE\_01419 A0A1R3XXU5\_MYCBO PknH2 protein  MAKKPEDRYASAGDLALAAHEALSDPDQDHAADILRRSQESTLPGTAAVTAQPPTMPTVT  PPPIQAAPTGQPSWAPNSGPMPASGPTPTPQYYQGGGWGAPPSGGPSPWAQTPRKTNPWP  LVAGAAAVVLVLVLGAIGIWIANRPDDNPKRNIATSPGTPTTTATTSLPATTTPTTAPAS  DPQTRLLSMLPSGYPTGTCKPTTPKPNSIWVNAVAMVDCGQNTNQGGPSRAIYGLFANPD  KLKQAFNDDIAAVELMNCPGEGPSPDGWHYNQTPDVTAGMIACGTYKNRPNVIWSNEAKL  TLSDVFGDPATIEDLHNWWAKYG |  |
| L12222 (SAMN15217201)  pknH | GCA\_013367655.1\_ASM1336765v1\_genomic  >CCEFDCB\_02241 PKNH\_MYCBO Serine/threonine-protein kinase PknH  MGPIPQPSAIRPGIPKAFDAVVARGMAKKPEDRYASAGDLALAAHEALSDPDQDHAADIL  RRSQESTLPAPPKPVPPPTMPATAMAPRQPPAPPVIPPGVQPAPKPSYTPPAQPGPAGQR  PGPTGQPSWAPNSGPMPASGPTPTPQYYQGGGWGAPPSGGPSPWAQTPRKTNPWPLVAGA  AAVVLVLVLGAIGIWIAIRPKPVQPPQPVAEERLSALLLNSSEVNAVMGSSSMQPGKPIT  SMDSSPVTVSLPDCQGALYTSQDPVYAGTGYTAINGLISSEPGDNYEHWVNQAVVAFPTA  DKARAFVQTSADKWKNCAGKTVTVTNKAKTYRWTFADVKGSPPTITVIDTQEGAEGWECQ  RAMSVANNVVVDVNACGYQITNQAGQIAAKIVDKVNKE |  |
| L12222  (SAMN15217201)  PknH2 | GCA\_013367655.1\_ASM1336765v1\_genomic  >CCEFDCB\_04052 A0A1R3XXU5\_MYCBO PknH2 protein  MLVLVLGAIGIWIANRPDDNPKRNIATSPGTPTTTATTSLPATTTPTTAPASDPQTRLLS  MLPSGYPTGTCKPTTPKPNSIWVNAVAMVDCGQNTNQGGPSRAIYGLFANPDKLKQAFND  DIAAVELMNCPGEGPSPDGWHYNQTPDVTAGMIACGTYKNRPNVIWSNEAKLTLSDVFGD  PATIEDLHNWWAKYG | NLA000017458(SAMN15217203) has the same protein coding sequence for its pknH2 |

Gene arrangement

There is a correlation between the strains with a higher scaffold number and those with pknH2 and tbD2 genes that were not found with the rest of the RD900 region (indeterminate).

The strains with a scaffold number > 90 differ in their gene arrangement than those with a smaller scaffold number – less complete gene arrangements are being produced in these cases.

Most complete?

Protein coding sequences – grouping

|  |  |
| --- | --- |
| Standard pknH sequence | K85  MAL010074  MAL010081  MAL010129  MAL010131  MAL020176  MAL020193  UT307 |
| Standard pknH2 sequence | MAL010081  MAL010131  MAL020176 |
| K85 pknH2 |  |
| MAL010118 pknH | MAL010129(has 2 pknHs – one same as MAL010118 and the other is the standard pknH sequence) |
| MAL010118 pknH2 | 25  MAL010129 |
| MAL020193 pknH2 | ATCC 25420 |
| 25 pknH | Three different pknHs for this strain |
| ATCC 25420 pknH |  |
| NLA00001748(SAMN14601854) pknH | NLA000017458(SAMN15217203) pknH |
| NLA00001748(SAMN14601854) pknH2 |  |
| L12222(SAMN15271203) pknH |  |
| L12222(SAMN15271203) pknH2 |  |
| L12222(SAMN15217201) pknH |  |
| L12222(SAMN15217201) pknH2 | NLA000017458(SAMN15217203) pknH2 |