

## Soybean phyllosphere microbial rhodopsin sequences

**D** or **E** mark the proton donor and acceptor sites

**L** mark position 105

>contig08968\_20 [877 - 1743]

RCTGGWRAPHDRWATSRVRLPGALVPRINATEAPMDIVTPFQYSLIYNAFSFTFATMAAATAFLWLSRSQVAPAYRTAL  
TISGLVTGIAAYHYRIFESWIAAYSFGNAVVTATPFAFNDAYRYVDWLLTVPLLLLELVLVMRLSREETSSKAVRLGS  
AAALMIVLGYPGEISADIPTRVLWGTLSLIPFVYIVWELFSGLGKSIADQPVAQAQGLVRKARLLTFASWGFYPIVYMAP  
YAGLTGGTVTTTTIQIGYTIADIVAKAGLGILIFLIAVRKSEAEADAHKGYAV

>SRR023845.12441

FGIGTLFKLDGSVTYLARYLDWSVTTPILLSSLVLLAFHEHGRTGEVGGHLTAIIIVLDVLMIAATGLVSSLTGPSFAKWV  
WYAW

>SRR023845.14845

LMIVLGYPGEISADIPTRVLWGTLSLIPFVYIVWE

>SRR023845.15385

SYFAMATGQGSIALPAGPEAADAARQFYFARYLDWTFTTPLLVGLLARTAMHSGMRRPAVVWGLVGSDDLIMIVTALAFG  
LSTVAWVKW

>SRR023845.23390

AEVLPRAVWGAASTIPFVYILYVLFVELSRSIVRQPVEVQHTIKMLRIMLVGLWGIYPIAYLFPILGFVGSDAFVLKQV  
GYSVAD

>SRR023845.42057

LLTFASWGFYPIVYMAPYAGLTGGTVTTTTIQIGYTIADIVAKAGLGILIFLIAVRKSE

>SRR023845.53547

VIAATGFAFNDAYRYVDWLLTVPLLLVLELVLVMRLSPEQTSSKTVRLGSAAALMIVLGYPGEIAADIPTRALWGTLSAI  
PFV

>SRR023845.54527

LISLAAPSLAKWIWYGWSCVAFLGVLYLLWGPLRSQALSARGAALGTAYAKNLIFLTVIWFYPIVFLVGPEGLKLITD  
PASVWAF

>SRR023845.79789

AVTISGLVTAIAAYHYLRIFESWNAAYTLKDGVIAATGFAFNDAYRYVDWLLTVPLLLVLELVLVMRLSPEQTSSKTVRL  
GS

>SRR023845.111891

YYRIFESWIAAYSFGNGVVTATPFAFNDAYRYVDWLLTVPLLLLELVLVMRLSREETSSKAVRLGSAAALMIVLGYPGE

>SRR023845.151708

RKARLLTFASWGFYPIVYMAPYAGLTGGTVTTTTIQIGYTIADIVAKAGLGILIFLIAVRKSEAEADAHKGYAV.CRDDR  
CAS

>SRR023845.159249

RTALTISGLVTGIAAYHYRIFESWIAAYSFGNGVVTATPFAFNDAYRYVDWLLTVPLLLLELVLVMRLSREETSSKA

>SRR023845.244868

LTPPILLSLVLLAFHEHGRTGEVGGHLTAIIVLDVLMIAATGLVSSLTGPSFAKWVWYAWSCVAFLGVLYLLWGPLRMLAKSRG

>SRR023845.278926

VWELFSGLGKSIADQPVAQAQLVRKARLLTFASWGFYPIVYMAPYAGLTGGTVTTTTIQIGYTIADIVAKAGLGILIFLI  
AVRK

>SRR023845.296499

LFVWWSRDPKGVPPQYEYSIAMMPIWSALAYLAMAMGQKTEVAGQTTHYARYADWVSTPLLLLALAFTAMFYVPKEK  
RSVT

>SRR023845.306873

AEVLPRAVWGAASTIPFVYILYVLFVELSRSIVRQPVEVQHTIKMLRIMLVGLWGIYPIAYLFPILGFVGSDAFVLKQV  
GYSVAD

>SRR023845.311061

YRYMDWAVTTPPLLLKASSMLRIRYHEAATAITVMLLADFFMVLTYIGEQQLTASGEIIASGKLLWGAISTVGYVVA

>SRR023845.329067

ARQFYFARYIDWFTTTPPLLLGLARTAMHSGMRRPAIVWGLIGSDLIMIVTALAFGLSTVAWIKWTWFAISCGAFLAVF  
Y

>SRR023845.363535

HTLLHASVPFIAATAYLAMAFGIGTLINSDGSLTYLARYVDSVTTPILLLSGLVLLAFHEHGEDG.GRWVSDLGHRAR

>SRR023845.384882

PAYRTALTISGLVTGIAAYHYRIFESWIAAYSFGNGVVTATPFAFNDAARYVDWLLTVPLLLIELVLMRLSREE

>SRR023845.408767

PAYRTALTISGLVTGIAAYHYRIFESWIAAYSFGNGVVTATPFAFNDAARYVDWLLTVPLLLIELVLMRLSREE

>SRR023845.412178

AATAFLWLSRSQVAPAYRTALTISGLVTGIAAYHYRIFESWIAAYSFGNGVVTATPFAFNDAARYVDWLLTVPLLLIE  
LVL

>SRR023845.436453

LIWFLYPIVFLVGPEGLKLITDPASVWAFLLIMDIIAKVGYAFYAAANLDKALAHQGDERGYRA.HDAPRPLAGASFPGR  
L.SH

>SRR023845.453555

AAALMIVLGYPGEIATDIPTRALWGTLSSIPFVYIVWELFRGLGESIEQQPAVARDLVRKARLLTFASWGFYPIVYMAP  
YCRD

>SRR023845.496732

TPFAFNDAARYVDWLLTVPLLLIELVLMRLSREETSSKAVRPGVGRRADDRAGLSGRDLGGYPDARPCGARCRRS

>SRR023845.497524

LAIYATGKKTQPSGHHTLLHASVPFIAATAYLAMAFGIGTLVNSDGSATYLARYVDSVTTPILLLSGLVLLAAHEHGKT  
GEV

>SRR023846.25381

AFLWLCRSQVAPAYRTAVTISGLVTAIAAYHYLRIFESWNESYTLKDGIIITATGFANDAYRYVDWLLTVPLLLIE

>SRR023846.42284

TLLHASVPIAATAYLAMAFGIGTLVNSDGSATYLARYVDWSVTTPILLLSGLVLLAFHEHGKTGEVGGYLTSVIVLDVLM

>SRR023846.45502

LMSVASLAIYATGKTHPSGHHTLLHASVPIAATAYLAMAFGIGTLINS DGSATYLARYVDWSVTTPILLLSGLVLLAFH

>SRR023846.50431

PSMAKWIWYGWSCVAF LGVLYLLWGPLRSLALS RGAALGMAYGKNLTF LTVIWF LYP I VFLIGPEGLKLITDPAS

>SRR023846.51795

IQIGYTIADIVAKAGLGILIFLI AVRKSEAEADAHKGYAV . CRDDRCASPSVIRTTTRVIRTTTRVIPSPCPIRAR . CRLSGLL

>SRR023846.66447

TATGFANDAYRYVDWLLTVPLLLIE LVLVMRLSPEQTSSKTVRLGFAAALMIVLGYPGEIAIDIPTRALWGTLSI

>SRR023846.143364

GEVGGHLTAIIVLDVLM IATGLVSSLTGPSFAKWVYAWSCVAF LGVLYLLWGPLRMLAKSRGMALGTAYNKNVAF LTVIWF LYP I

>SRR023846.143499

PRVLWGTLS SIPFVYI VWELF SGLGKSIADQPVA AQGLVRKARLLTFASWGFYPIVYMAPYAGLTGGTVTTTIQIGYTIADIV

>SRR023846.149776

GFTGGAATTAVQVGYTVADIVAKAVFGVLIYMI AVRKSAADGYVSGTLAERNMPA . THIREVRTTSTGAIGSSGSF . SLERK

>SRR023846.160842

TLLAADLFMIITGYIGEQQLGANGEILVGPKLIWGAVSTIGYIVVYVLYKIWQDFSNQSYVKEPERQAYKVMGLLVVTMWGVYPI

>SRR023846.164342

MDIVTPFQYSLIYNAFSFTFATMAAATAFLWLRSQVAPAYRTALTISGLVTGIAAYHYRIFESWIAAYSFGNGVVTA TPFAN

>SRR023846.178011

LINS DGSATYLARYVDWSVTTPILLLSGLVLLAFHEHGKTGEVGGYLTSVIVLDVLMIVTGLISSLAAPSLAKWIWYGW

>SRR023846.211212

GVVTATPFANDAYRYVDWLLTVPLLLIE LVLVMRLSREETSSKAVRLGSAAALMIVLGYPGEISADIPTRVLWGT

>SRR023846.243929

LTVPLLLIELVLVMRLSREETSSKAVRLGSAAALMIVLGYPGEISADIPTRVLWGTLS SIPFVYI VWELF SGLGKSI

>SRR023846.269836

DAYRYVDWLLTVPLLLIEFLVLMRLSREETSSKAVRLGSAAALMIVLGYPGEISADIPTRVLWGTLS SIPFVYIV

>SRR023846.288987

QRARLLTFASWGFYPIVYMAPYAGLTGANVTTVIQLGYTIADIVAKAGVGVLI FMISSRKSQAEGHVAGVHAA.PT

>SRR023846.327332

ASVPFIAATAYLAMAFGIGTLINSDGSLTYLARYVDWSVTTPILL SGLVLLAFHEHGKTGEVGGYLTSVIVLDVLMVVT  
G

>SRR023846.335157

TATGFADFNDAYRYVDWLLTVPLLLIEFLVLMRLSPEQTSSKTVRLGFAAALMIVLGYPGEIAIDIPTRALWGTLS SI

>SRR023846.383390

PFIAATAYLAMAFGIGTLINSDGSLTYLARYVDWSVTTPILL SGLVLLAFHEHGKTGEVGGYLTSVIVLDVLMVVTGLI  
S

>SRR023846.384866

AKWIWYGWSCVAFLGVLYLLWGPLRSQALS RGAALGTAYAKNLI FLTVIWFYPIVFLVGP EGLKLITDPASVWAFLIM  
DIIA

>SRR023846.385744

PRVLWGTLS SIPFVYIVWELFSGLGKSIADQPVA AQGLVRKARLLTFASWGFYPIVYMAPYAGLTGGTVTTTIQIGYTI  
ADIV

>SRR023846.400375

ARQFYFARYIDWTF TTPILLVGLARTAMHSGMRRPAVVWGLVGS DLIMIVTALAFGLSTVAWVKWTF AISCGAFLAVF  
YGIFVQ

>SRR023846.404753

AATAYLAMAFGIGTLFKLDGSV TYLARYLDWSVTTPILL SSVLLAFHEHGRTGEVGGHLTAIIVLDVLM IATGLV

>SRR023846.413518

AATAFLWLSRSQVAPAYRTALTISGLVTGIAAYHYRIFESWIAAYSFGNGVVTATPFAFNDAYRYVDWLLTVPLLLIE

>SRR023846.439417

PRVLWGTLS SIPFVYIVWELFSGLGKSIADQPVA AQGLVRKARLLTFASWGFYPIVYMAPYAGLTGGTVTTTIQIGYTI  
ADIV

>SRR023846.475949

AFLWL CRSQVAPAYRTAVTISGLVTAIAAYHYLRIFESWNESYTLKDGIIITATGFADFNDAYRYVDWLLTVPLLLIE

>SRR023846.481799

FVYIVWELFSGLGKSIADQPVA AQGLVRKARLLTFASWGFYPIVYMAPYAGLTGGTVTTTIQIGYTIADIVAKAGLGIL  
I