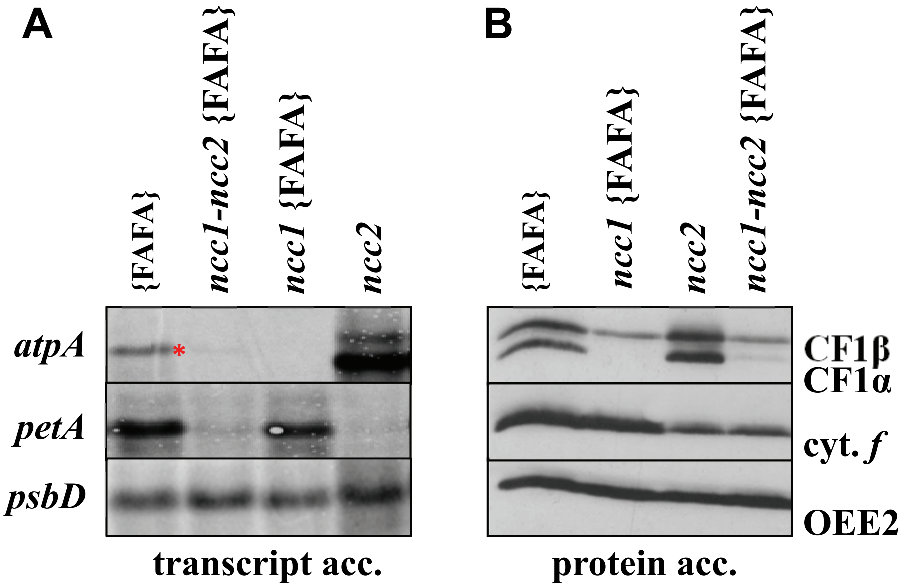


Supplemental Figure 1: Phenotype of the double mutant *ncc1 ncc2* {FAFA}, descendant of the cross *ncc1 mt*+ {FAFA} x *ncc2 mt-*, whose genome was sequenced.

A. *petA* and *atpA* transcript accumulation in the parental strains *ncc2* and *ncc1* {FAFA}, in a {FAFA} transformant as a control and in the double mutant *ncc1 ncc2* {FAFA}, whose genome was sequenced to identify the mutations. *psbD* serves as loading control. The red asterisk points to the *FAFA* chimeric transcript, which migrates between the mono- and the di-cistronic *atpA* transcripts.

B. Accumulation of CF1-subunits α and β (detected using an antibody raised against the whole CF1 complex (Lemaire and Wollman, 1989)) and of cytochrome *f* (OEE2 serves as loading control) in the same strains.

```
NCL2        1
NCL36       1 MFQFAHG--------SAASGYGLGSRYNA----PASANTGL---NLCPCGIPQTQPRAARTAGRLGRLAPSAT-----------------------TGGVARGLACSSATSYPAAAVLSGGGPHLAAAAPSATG AGG---
NCL37       1 MLQFAHP--------SPANGHGLGSSYTS---APASPLIWH---RMCPCGIPQTQPRAVRTAKRL---GPSAT-----------------------TGGVARGLACSSATSGPAAAVLSGGGPRLAAAAPAVTG AGG---
NCL38       1 MLQFAHP--------SPANGHGLGSSYTS---APASPLIWH---RMCPCGIPQTQPRAVRTAKRL---GPSAT-----------------------TGGVARGLACSSATSGPAAAVLSGGGPLLAAAAPAVTG AGG---
NCL23       1 ---------------------------------------------MCPCGIPQTQPRAARTTGRL---GPSAT-----------------------AGRVARGLACSSATSGPAAAVLSSSSPHLAAATPAATW AGG---
NCC1        1 MLQFAHGSA------VG-----LR-RSYN---APASAHIGL---ELWPCGIRLTQPRAARTAGRL---GPSAT-----------------------AGRVARGLACSSATSGPAAAVLSGGGPRLAATAPAATG AGG---
NCL24       1 MLQFTRASA------AAASGHDLGSRYTS---APASALVWH---RVCPCGIPLTQPRAARTAGRL---GPSAT-----------------------AGGVARGLACSSATSGPAAAVLSSSSPHLAAATPAATW AGG---
NCL22       1 MLQFTRASA------AAASGHDLGSRYTS---APASALVWH---RVCPCGIPLTQPRVARTTGRL---GPSAT-----------------------AGGVARGLACSSATSGPAAAVLSSSSPHLAAATPAATW AGG---
NCL25       1 ML-LFTRAS------AAASGHDLGCRYTS---APASALIWH---RVCPSGIPQS---AARTAGRL---GPSAT-----------------------AGGLARGLACSSATSGPAAAVLSGGGPRLAAATPAATG AGG---
NCL17       1 MLQ--SGRG------TAASRHGLGSSYN----ASASVRTWH---RVGPCGIPLTQPRAARTAGRL---GPSAT-----------------------AGGVARGLACSSATSGPAAAVLSGGGPRLAAAAPAATG AGG---
NCL21       1 MLQFAHGSAV-----------GLRRSYN----APASAHIGL---ELCPFGIRLTQPRAARTAGRL---GPSAT-----------------------AGGVARGLACSSATSGPAAAVLSDGGPLLPATAPAATG AGG---
NCL19       1 MLQ--SGRG------TAASRHGLGSSYN----APASARTWH---RVGPCGIPLTQPRAARTAGRL---GPSAT-----------------------AGGVARGLACSSATSGPAAAVLSEGGPLLPATAPAATG AGG---
NCL20       1 MLQ--SGRG------TAASRHGLGSSYN----APASARTWH---RVGPCGIPLTQPRAARTAGRL---GPSAT-----------------------AGGVARGLACSSATSGPAAAVLSEGGPLLPATAPAATG AGG---
NCL15       1
NCL35       1
NCL14       1 ML--RAGRG------SAASGHGLGSSYN----APASARTWH---RMCPCGTPLTQPRAARTAGRL---GPSAT-----------------------AGGLARGLTCSSATSGPAAAVLSGGFPRLAATAPAATG AGG---
NCL1        1 MLQFAHG--------SAASGHGLGSNYN----APASARTWH---RMCPYGIPLTQPRAARTAGRL---GPSAT-----------------------TSGVARGLA--SATSGPAAAVLSGGGPRLATTAPSATG AGG---
NCL16       1                                                  MLQVAQPRAARTAGCL---GPSAT-----------------------AGGVARGLA--SATSGPAAAVLSGGGPLLAATAPAATG▼SGG---
NCL30       1 MLAQL-SRALILASATGARGQALGSLSSAIRNRCLEHHQHLTTTTEDGC-VVRSAPLGSRCVGLS---HPAAAAGSEAPAAADHSSDLLQLPLPLPASHLAYTTSYSSVISGPAAAGLSGGGPRLAAAAAPAMF TPGAAA
NCL34       1 MLQFARASA-------AASGHGLGSSYTS---APVSAHTGLT---LRPCGIPLTQPRAARTAGRL---QPLAT-----------------------AGGVARGLACSSATSGPVAAVLSGGGPCLATAAPAVTG AGG---
NCL33       1 MLQFARASA-------AASGHGLGSSYTS---APASAHTGLT---LRPCGIPLTQPRA---AGRL---HPPAT-----------------------AGGVARGLACSSATSGPAAAVLSGGGPCLATAASAVTG AGG---
NCL28       1 MLQFARASA-------AASGHGLGSSYTS---APASAHTGLT---LRPCGIPLTQPRASRTAGGL---HPSAT-----------------------AGGVARGLACSSATSGPAAAVLSGCGPRLAAAAPAATG AGG---
NCL29       1 MLQFARGSA-------AASRHGLGSSHTS---APVSARTGLT---LRPCGIPRTQPRA---AGRL---HPPAT-----------------------AGGVARGLACSSATSGPAAAVLSGGGPCLATAASAVTG AGG---
NCL27       1 MLQFARGSA-------AASGHGLGSSYTS---APVSAHTGLT---LRPCGIPLTQPRA---AGRL---HPSAT-----------------------AGGVARGLACSSATSGPVAAVLSGGGPCLATAAPAVTG AGG---
NCL31       1 MLQFARGSA-------AASRHGLGSSHTS---APVSAHTGLT---LRPCGIPLTQPRA---AGRL---HPSAT-----------------------ARGVARGLACSSATSGPVAAVLSGGGPCLATAAPAVTG AGG---
NCL32       1 MLQFARASA-------TASRHGLGSSHTS---APVSAHTGLT---LRPCGIPLTQPRA---AGRL---HPSAT-----------------------AGGVALGLACSSATSGPAAAVLSGGGPCLATAAPAVTG AGG---
NCL26       1 ML-TMR----------------------------------------------------------------------------------------------------------TRL------FGNTRNSHGCFTV S-----
NCL3        1 M----RTGLL------------------------------------------------------------------------------------------------GR----PPL------SAQSNSGHGY--G Y-----
NCC2        1 M----RT---------------------------------------------------------------------------------------------------SR----PPL------SAYSNSRHGY--G Y-----
NCL5        1 MLAMRT----------------------------------------------------------------------------------------------------SR----PPL------SAYSNSRHGY--G Y-----
NCL6        1
NCL13       1 MLVMRA----------------------------------------------------------------------------------------------------S---SGLVAGSFSGTCARSHSGHGY--C N-----
NCL10       1 MLALRV----------------------------------------------------------------------------------------------------V---AGSPA------GAHRHSVHGC--S H-----
NCL7        1 MLALRV----------------------------------------------------------------------------------------------------V---AGSPA------GAHRHSVHGC--S Q-----
NCL8        1
NCL9        1 MLALRV----------------------------------------------------------------------------------------------------V---AGSPA------GAHRHSVHGC--S Q-----
NCL11       1 MLVMRA----------------------------------------------------------------------------------------------------SGRFAGSPA------GAHSHLSGHG--S H-----
NCL12       1
consensus   1 ml                                                                   p a                           va     ss  sgp aa ls  gp l       t    g
```

Supplemental fig. 2

```
NCL2        1
NCL36     100 --GLRQLLPLAAWARQTAQAPATASPAAAGARSCLLQPGRRAQQQVLAAAGR---YGSDRSSDS---SS-------------SSGRGYI------SGSRS----SSDGGGRVGRGRAGGGRWAGGGG-RGTGRTGGG-GG
NCL37      98 --GLQQVLPLAAWARQAAPAPTAAAPAAAGARSHLLQPGRRAQQQVLAAAGR---YGSGGTSDSSSDSS-------------SSRRGGI------SGSRS----SSDGGGRAGRGRADRGRWAGGGG-RGSGRTGGG-G-
NCL38      98 --GLQQVLPLAAWARQAAPAPTAAAPAAAGARSYLLQPGRRAQQQVLAAAGR---YGSGGTSDRSSDSS-------------SSRRGGI------SGSRS----SSDGGERAGRGRADGGRWAGGGG-RGSGRAGGG-G-
NCL23      67 --GLRQLLPLAAWARQAAPAPATAAPAAAGARSCLLQPGRRAQQQVLAAAGR---YGSGSTSHRSIVSS-SDSSSSG-H-------GGI------GGSRS----SSDGGGRAGRGRAGRGHFASVGG-RGSGRTGGG--G
NCC1       94 --GLRQLLPLAAWARQAAPAPATAAPAAAGARSCLLQPGRRAQQQVLAAAGR---YGSGGTSHRSSGRS-SDNSSSG-H-------GGI------SGSRS----SSDGGGRAGRGRADGGHWAGGSG-RGSGRTGGG--G
NCL24     100 --GLRQLLPLAAWARQAAPAPATAAAAAAGARSCLLQPGRRAQQQVLAAAGR---YGSGGTSDKSSGRS-SDNSSSG-H-------GGI------SGSRS----SRDGGGRAGRGRADGGRWAGGGG-RGSGRTGGG--G
NCL22     100 --GLRQLLPLAAWARQAAPAPATAAPAAAGARSCLLQPGRRAQQQVLAAAGR---YGSGGTSDRSSGRS-SDNSSSG-R-------GGI------SGSRS----SRDGRGRAGRGRADGGRWAGGGG------------R
NCL25      96 --GLRQLLPLAAWARQAAPAPATAAAAAAGARSCLLQPGRRAQQQVLAAGGR---YGSGGTSHRSIDR----SSSSG-H-------GGI------GSSRS----SSDAGGRAGGGRAGGGRWAGG-------RTGGA--G
NCL17      97 --GLRQLLPLAAWARQAAPAPATAALAAAGARSCLLQPGRRAQQQVLAAAGR---YGSGSSSGRGSD-----NSSSG-H-------GGI------SGSRS----SSDGGGRAGRGRADGGRWAGGSG-RGSGRTGGG--G
NCL21      94 --GLRQLLPLAAWARQAAPAPATAAPAAAGARSCLLQPGRRAQQQVLAAAGR---YGSGGTIDRSSGR----SSSSG-H-------GGI------GSSRS----SSDAGGRAGGGRAGGGRWAGG-------RTGGA--G
NCL19      97 --GLRQLLPLAAWARQAAPAPAPAAPAAAGARSCLLQPGRRAQQQVLAAGGR---YGSGGTSHRSIDR----SSSSG-H-------GGI------GSSRS----SSDAGGRAGGGRAGGGRWAGG-------RTGGA--G
NCL20      97 --GLRQLLPLAAWARQAAPAPAPAAPAAAGARSCLLQPGRRAQQQVLAAGGR---YGSGGTSHRSIDR----SSSSG-H-------GGI------GSSRS----SSDAGGRAGGGRAGGGRWAGG-------RTGGA--G
NCL15       1
NCL35       1
NCL14      97 --GLRQLLPLAAWARQAAPAPATAAPAAAGARSCLLQPGRRAQQQVLAAAGR---YGSGGTSHRSSGRS-SDSTSDS-SIDSSSGRGSD------SSSNNDRSRGGHGADRAGRGRAGGGRWAGGRTGGGSGRTGGG---
NCL1       95 --GLRQLLPLAAWARQAAPAPATAAPAAAGARSCLLQPGRRAQQQVLAAAWR---YGSGGTSYTSSDRS------------------SD------SSNNNDKGRGGHVEDRAGRGRAGGGRWAGGRTGGGSGRSGGG-GG
NCL16      61 --GLRQLLPLTAWARQAAAAPATAAPAAAGARSCLLQPGRRAQQQVLAAAGR---YGSGSTSDKSSGRS------------------SG------SSNNNDKGRGGHVEDRTGRGRAGGGRWAGGRTGGGGSRS------
NCL30     136 AAAYGSLPPPLGW---------------------------HAQPQVAAAGVM---HGSGTTGSASSGSS-GSS-----S-------------------SS----SSSGEGDGGNGRSSGG-----------GRGSGRTGG
NCL34      99 --GLLQLLPLAAWARQAAPAPATAAPAAAGARSCLLQPGRRAQHQVLAAAGR---YGSGGAIDRSSGMS-SSS--------SSSGRDDI------SGSRS----SSDGGGRAGRGRAGGGRWNGGRRG---------SGR
NCL33      96 --GLRQLLPLAAWARQAAPAPATGAPAAAGARSCRLQPGRRAQHQVLAAAGR---YGSGSSSGRG-------------SDNSSSGHGGI------SGSRS----SSDGGGRAGRGWAGGGHWADGRTGGGGSMSAGD---
NCL28      99 --GLRQLLPLAAWARQAAPAPAAVAPAAAGARSCLLQPGRRAQQQVLAAAGR---YGSGGIIDRSSGMS-SSSSSSSSSSSSSSGRDDI------SGSRS----SSDGGGKAGRGRAGGGRWNGGRRG---------SGR
NCL29      96 --GLWQLLPLAAWARQAAPAPATAAPAAAGARSCLLQPGRRAQQQVLAAVGR---YGSGGIIDRSSGMS-SSSSSS--SSSSSSGRDDI------SGSRS----SSDGGGKAGRGRAGGGRWNGGRRG---------SGR
NCL27      96 --GLRQLLPLAAWARQAAPAPATAAPAAAGARSCLLQPGRRAQQQVLVAVGR---YGSGGIIDRSSGRS-SDS-------------------------SS----SSDGGGRAGRGRAGGGRWAGGRTGGAGGRSSG----
NCL31      96 --GLRQLLPLAAWARQAAPAPATAAPAAAGARSCLLQPGRRAQHQVLVAAGR---YGSGGIIDRSSGMS-SSS-------------------------RS----SSDGGERAGRGRAGGGRWARGRTGGAGGRSSGGDGR
NCL32      96 --GLRQLLPLAAWARQAAPAPATGAPAAAGARSCRLQPGRRAQQQVLAAAGR---YGSGS--------------------------------------RS----SSDGGGRAG------------------------RGR
NCL26      23 ----------HMLAVSKGPRPAGALL-----------PKTAG---------PSS----------------SVPRPSQRAGSAVAVLGGGSEL--FLVPVASRQ---LVQAAAAGGGAAGGEGGWGRSGGLGGGHAGRDSS
NCL3       24 ----------GYAAVSQGPRPASRVL----------LPVRSGTTVPRAFAGPSAARVAAPAPALR---S------SSTWSGSEAVLGGIRRLAPATPPTSRRQ---VVQTAAAGSGAAAGGANSGG-RGSGRG-YRGSGT
NCC2       21 ----------GSAAVSQGPRPASRGL----------LPVRSGTTVPRAFASPSSAYSARPVLLES---RSLSPR-SGSVPSVLAGSGGRRPIAPAVPPPSSRQ---LEQAAAGGGGASSS-----S-RGEGR--MGGRGR
NCL5       24 ----------GYVAVSQGPRPASRVL----------LPVRSGTTVPRAFASPSS---ARPALLES---GSLSPRRSGSVPTVLAVGSGRRPIAPAVPPPSSRQ---LEQAAAGGGGAGSS-----S-RGEGR--MGGRGR
NCL6        1
NCL13      31 ----------SSSARLQAPQPACALR----------RLFGR------SDSDPSSPRAAAATSVPAHNSSSSSPL----RVSTAAVLGGS------SSSRLGRS---VVQVAAGRGSSGGRAGGG--------------S-
NCL10      25 ----------SLSARLQAPPPACAFR----------RLFVYGMHRTAADTGPSSPRAAAATSVPALNSSSSSPL----RVSTAVVLCGS------SSSRLGRP---IVQAAAGRGSGGGRAGDG--------------SR
NCL7       25 ----------SLSARLQAPPPACAFR----------RLFVYGMHRTAADTGPSSPRAAAATSVPALNSSSSSPL----RVSTAVVLCGS------SSSRLGRP---IVQAAAGRGSGGGRAGDG--------------SR
NCL8        1 ----------------------------------------------------------------------------------------------------------------GRGSGGGRAGDG--------------SR
NCL9       25 ----------SLSARLQAPPPACAFR----------RLFVYGMHRTAADTGPSSPRAAAATSVPALNSSSSSPL----RVSTAVVLCGS------SSSRLGRP---IVQAAAGRGSGGGRAGDG--------------SR
NCL11      28 ----------RSPACLQAPRPVVTAM----------PAARGIAPSRAPASDPSSARAAATASLRARY--SPFSL----RGLMVAMRGGS------SSTRLARP---VVQAAAGRGSGGGGRGRGGGMAGG------GRGG
NCL12       1                                                                      MPGRL----QATLTATRGFL------AQPQVQRR---FVQTAGGRGSGGGGGGRTGG------------GG
consensus 141   g   l p aawar  ap pa aa            pgrraq qv aaag    ygsg                                       rs          g agrg aggg   gg
```

Supplemental Fig. 2A (continued)

```
                                                                                                                           OPR_A
                                                                                                      |----------------------------------------|      |----
                                                                                                         ↓3 ↓6                                           ↓3
NCL2        1                                             MSLL--APSRPAARVSTRRKAATAAIRASIIATLTPALLPLVPRIRNPFGCSIPLWVLAKAGA---ASDERVESQLAPVLLQRLGD-PVLLKHAAPQ
NCL36     207 --GGGGGRARDPADDMTMCRTLEALQAVIDRRLAVWSERQDVSTMSAAFGRCGK▼LESACAGGPAATAAARAGIMAALAPALLPLVPRIRQPVGCSLPLWALARAGA---ASDGHVESQLAPALLQRLVD-PVLLEAAKPQ
NCL37     207 --GGGGGRPRDPSDDMTMCRNLEALQAVIGQRLSMWSGRQDVSTMSVAFNLCGK▼LESACAGGPAATAAARAGIMAALACAYLPLVPRIRHPADCSIPLLALAKTGA---ASDGRVESQLAPALLQRLVD-PVLLEAAKPQ
NCL38     207 --GGGGGRARDPADDMTMCRNLEALQAVIGQRLSVWSERQDVSTMSVAFNLCGK▼LESACAGGPAATAAARAGIMAALACAYLPLVPRIRHPADCSIPLLALAKTGA---ASDGRVESQLAPALLQRLVD-PVLLEAAKPQ
NCL23     180 ---GGGGRARDPADDMTVCRTLEELQGVIDQRLAVWCERQDVSSMSAAFNLCGK▼LESARAGGPAATAAARAGIIAALAPALLPLVPRIRQPAGCSLPLLALSKAGA---ASEGRVESQLASALLQRLVD-PVLLDSATPQ
NCC1      207 GRSGGGGRPRDPEDVMIGYATLEELQGVIDQRLAVWCERQDVSTMSAAFGRCGK▼LESARAGGPTATAAARANIIAALAPALLPLVPRIRQPAGCSLPLLALAKAGA---ASDGRVESQLAPALLQRLVD-PVLLDSATPQ
NCL24     213 GRSGGGGRPRDPEDVMIGYATLEELQGVIDQRLAVWCERQDVSTMSAAFGRCGK▼LESARAGGPAATAAARAGIIAALAPALLPLVPRIRQPAGCSIPLLALSKAGA---ASDGRVESQLAPALLQRLVD-PVLLDSATPQ
NCL22     204 GRSGGGGRPRDPEDLMIGYATLEELQGVIDQRLAVWCERQDVSTMSAAFGRCGK▼LESARAGGPTATAAARAGIIAALAPALLPLVPRIRKPVDCSIPLLALAKAGA---ASDGRVESQLAPALLQRLVD-PVLLDGATPQ
NCL25     200 GRSGGGGRARDPADDMTVCRTLEDLQAVLDQRLAVWCERQDVSTMSAAFGRCGK▼LEPARAGGPAATAAARAGIIAALAPALLPLVSRIRQPAGCSIPLWALAKAGA---AIDGRVESELAPALLQRLVD-PVLLDSAIPI
NCL17     206 GRSGGGGRPRDPEDVMIGYATLEELQGVIDQRLAVWCERQDVSAMSAAFGRCGK▼LESARAGGPTATAAARAGIIAALAPALLPLVPRIRQPASCSIPLLALSKAGA---ASDGRVESELAPALLQRLVD-PVLLESATPQ
NCL21     198 DRSGGGRRARDPADDMTVCRTLEQLQGVIDQRLAVWCERQDVIAMSAAFGRCGK▼LESARAGGPAATAAARAGIIAALAPALLPLVPRIRQPAGCSLPLWALAKAGAA*AASDGRVESQLAPALLQRLVD-PVLLDSATPQ
NCL19     201 DRSGGGRRARDPADDMTVCRTLEQLQGVIDQRLAVWCERQDVIAMSAAFGRCGK▼LEPARAGGPTATAAARAGIIAALAPALLPLVPRIRQPASCSLPLWALAKAGA---ASDGRVESQLAPALLQRLVD-PVLLDSAIPI
NCL20     201 DRSGGGRRARDPADDMTVCRTLEQLQGVIDQRLAVWCERQDVIAMSAAFGRCGK▼LESAHASGPAATAAARAGIIAALAPALLPLVPRIRQPAGCSLPLWALAKAGA---AIDGRVESELAPALLQRLVD-PVLLDSATPQ
NCL15       1                MIGYATLEALQAVIDRRLAVWSERQDVSTMSAAFGRCGK▼--SARAGGPTATAAARAGIMAALAPALLPLVPRIRQPADCSLPLWALAKAGA---ASDGRVESQLAPALLQRLVD-PVLLKTATPQ
NCL35       1                                             MSAAFNLCGK▼LESARAGGPAATAAARAGIMAALAPALLPLVPRIRQPAGCSIPLWALAKAGA---ASDGRVASQLAPALLQRLLD-PVLLEAAKPQ
NCL14     221 --GGGGGRARDPADDMTVCRTLEDLHDVIDKRLAVWSERQDVMTMSAAFNLCGK▼LESARGGGPAATAAARAAIVAALAPALLPLVPRIRQPAGCSLPLWALAKAGA---ASDGRVESQLAPALLQRLVD-PVLLKAAKPQ
NCL1      205 GGGGGGGRARDPADDMTVCRTLEDLQAVIDRRLAVWSERQDVSTMSAAFNLCGK▼LESARTSGPAPTAAARFRIVAALSPALLPLVPRIRQPAGCSLPLWALAKAGA---AIGGRVESQLAPALLQRLAD-PVLLKTATPQ
NCL16     166 --GGGGGKARDPADDMSVCRTLEDLHDVIDRRLVVWSERQDVMTMSAAFNLCGK▼LDSARAGGPAATAAARAGIVAALAPALLPLVPRIRQPAGCSIPLLALAKAGA---AIDGRVESQLAPALLQRLAD-PVLLKTATPQ
NCL30     206 NGGGGGGRARDPTDDMTVCRTLEDLQAVIDQRLAVWCERQDVSAMSAALGRCGK▼LESARA-GPAATAAARAGIIAALAPALLPLVPRISKPVDCSIPLWALAKAGA---ASDGRVESQLAPALLQRLVD-PVLLDSATPQ
NCL34     206 SSGGGGGRARDSADDMTMCRTLEELQAVIDQRLAVWCERQDVSTMSAAFNLCGK▼LESSHAGGPATTAAARAGIMAALAPALLPLVPRIRQPASCSIPLWALAKAGA---ASDGRVESQLAPALLQRLVD-PVLLDSATPQ
NCL33     205 -RSGGGGRARDPADDMTVCRTLEQLQGVIDQRLAVWCERQDVSTMSAAFNLCGK▼LESAHAGGPAATAAARAGIIAALAPALLPLVPRIRQPASCSLPLLALAKAGA---ASDGRVESQLASALLQRLVD-PVLLETATPQ
NCL28     214 SSGGGGGRARDSADDMTMCRTLEELQAIIDQRLAVWSVRQDVSTMSAAFNLCGK▼LESARAGGPAATAAARAGIMAALAPALLPLVPRIRQPAGCSIPLWALAKAGA---ASDGRVESQLAPALLQRLVD-PVLLKTATPQ
NCL29     209 SSGGGGGRARDSADDMTMCRTLEELQAISDQRLAVWSVRQDVSTMSAAFNLCGK▼LESARAGGPAATAAARAGIMAALAPALLPLVPRIRQPAGCSIPLWALAKAGA---ASDGRVESQLAPALLQRLVD-PVLLDSATPQ
NCL27     197 ---AGGVRARDPADDMTVCRTLEDLQAVIDQRLAVWCERQDVMTMSAAFNLCGK▼LESARAGGPAATAAARAGIMATLAPALLPLVPRIRNPASCSLPLWALAKAGA---ASDGRVESQVAPALLQRLVD-PVLLDSATPQ
NCL31     201 TGGGGGGRARDPADDMTVCRTLEELPAVIDQRLAVWSERQDVMTMSAAFNLCGK▼LESSHA-GPAATAAARAGISAALAPALLPLVPRIRQPASCSIPLWALAKAGA---ASDGRVESQLAPALLQRLVN-PVLLDSATPQ
NCL32     165 AGGGGGGRARDPADDMTVCRTLDELQVVIDQRLAVWCERQDVMTMSAAFNLCGK▼LESSHA-GPAATAAARAGIMAALAPALLPLVPRIRQPASCSIPLWALAKAGA---ASDGRVESQLAPALLQRLVN-PVLLGMAKPQ
NCL26     112 GRGRGRGQGQDST---PNDGTLEELQAAVTDQLPGWVERHDTAAISNAFRKAVQ▼LRSG-DGGAATAARIRSGIIAGLAAGYLRLVPRIRKAIECRIPLWAMAKAGIS--GK--GKEAELTAALLQRLVTDSRLLGAATPM
NCL3      130 GRGGGSRGRGSPP---ADNQTLEELRAVVNGQLPGWVARKDNASIWRAFNKAAK▼LRSG-DGGAATAARIRSGIIADLSAAYLPLGPAVCAAVECSIPLWAMAKVGIS--GQ--GKEAQLAAALLERLLD-PAVLSTATPM
NCC2      126 GRGGFRGRGGTAA---SDDETLEELQAAVTCQLPGWVEREDTAAISNAFRKAVQ▼YRSG-NGGAATAARIRSGIIADLSAAYLPLVPRIRKPFDCRMPLWALGKAGIG--SK--GPEVQLANALLQRLLD-PAVIAAATAM
NCL5      127 GRGGFRGRGGTAA---SDDEMLEELQAAVTCQLPGWVEREDTAAISNAFRKAVQ▼LRSA-DGGAAMAARISSGIIADLSAAYLPLVPRIRKAIDCRMPLWALGKAGIG--SK--GPEAQLANALLQRLLD-PAVIAAATPM
NCL6        1                MTVCQTLEELQALISQRLSVWKGRKDVTTMFAAFNLCGR▼LKSR-AG-AAATAVARNSIIGALAPAYLPLVPRIRDAKNCSIPLLALGKAGVA--GGGGGTEAQLAAALLERLVD-PEVLVDAQPQ
NCL13     117 ASGGGRGSGRDPADDMSVCQNLEDLQTFITRRLPVWEGKGDVITMCAAFNKCGK▼LESR-AG-ANQTVVARSSIIAALAPAYLPLVPRMRMAVDCSLPLWALGKAGVS--GD---TEAQLAAALLERLVV-PEVMGTAKPQ
NCL10     118 ASGRSRGSGRDPADDMTVCRTLEDLQAIISQRLSVWEGRKDFITMCAAFNLCGK▼LESR-AG-ASQTAVARSSIIGMLAPAYLPLVPRIRKAVDCSIVLWALGKAGIS--GG-SAMEAQLIAALLERLLD-SEVLGTAKPQ
NCL7      118 ASGRSRGSGRDPADDMTVCRTLEELQAIISQRLSVWEGRKDFITMCAAFNLCGK▼LESR-AG-ASQTAVARSSIIGMLAPAYLPLVPRIRNAKNCSIPLWALGKAGLV--GG-G-TEAQLAAALLERLVD-PEVLGKAEPQ
NCL8       15 ASGRSRGSGRDPADDMTVCRTLEELQAIISQRLSVWEGRKDFITMCAAFNLCGK▼LESR-AG-ASQTAVARSSIIGMLAPAYLPLVPRIRNAKNCSIPLWALGKAGLV--GG-G-TEAQLAAALLERLVD-PEVLGKAEPQ
NCL9      118 ASGRSRGSGRDPADDMTVCRTLEELQAIISQRLSVWEGRKDFITMCAAFNLCGK▼LESR-AG-ASQTAVARSSIIGMLAPAYLPLVPRIRNAKNCSIPLWALGKAGLV--GG-G-TEAQLAAALLERLVD-PEVLGKAEPQ
NCL11     127 SGGSGRGSGRDPADDMTVCRTLEELQATISQRLPVWEGRKDFITMCAAFNLCGR▼LESS-GGPAAATAAARSSIIGTLAPAYLPLVPRIRKARDCSIPLWALGKAGLF---G-GGTEAQLAAALLERLVD-PEVLAKAKPQ
NCL12      47 ASGGGRGGARDPADDMDVCQNLEDLQALISQRLSVWERKKDVVTMCAAFSKCGN▼LESR-SAGAAATAAARTGILAVLAPAYLPLVPRIRNAMKCSIPLWALGKAGVS--SS-GGTEAQLVAALLERLLD-SEVLGTARPQ
consensus 281   gggggrardpaddmtmcrtleelqavi qrl vw erqdv tmsaaf  cgk lesa aggpaataaaragIiaaLapalLpLvprirnp  CsipLwalakaGa   asdgrvesqlapaLLqRLvd pvll  A pq
```

Supplemental Fig. 2A (continued)

```
                       OPR_B                                                      OPR_C                                   OPR_D
              ----------------------------------|                |------------------------------------||---------------------------------------|     |----
              ↓6                                                   ↓3 ↓6                                 ↓3 ↓6                                         ↓3
NCL2       92 SLANALYALGKLREDQQHRG-SGWDPTA-SAHLT-ALARAVASRLRGAWGHGFNAQDTSNSLWACAKLGYRG-SELSLPLAEAAAALVRSMRPQELPNTLWALKALSCTGPMF*AVLETLCGETLRRLRAPELAAAFKPQ
NCL36     341 ELSNALYALGKLREQQQQRG-SGWDPTS-SPHLN-ALTGAVASRLRAAEGHGFNAQNVSNSLWACAKLGYRD-SDLLLPLVEAAAALARGMNAQDVANSLWALEVLGCTGPAFRAGLETLCGAALRQLRTPKEAEAFKPQ
NCL37     341 ELSNALYALGKLREQQQQRG-SGWDPTS-SPHLN-ALTGAVASRLRAAEGHGFKPQELSNSLWVCAKLGYRD-SDLLLPLVEAAAALARGMNAQDVTNNMWALEALGCTGPEFRAVLETLCGAALRQLRTPKEAEAFKPQ
NCL38     341 ELSNALYALGKLREQQQQRG-SGWDPTS-SPHLN-ALTGAVASRLRAAEGHGFNAQNVSNSLWACAKLGYRD-SDLLLPLVEAAAALARDMNAQGVANSLWALEALGCTGPAFLAGLETLCGAALRQLRTPKEAEAFKPQ
NCL23     313 ALANALYALGKLREDQQQRG-SGWDPTS-SPHLMNALTGAVASRLRAVSGHGFEPQHVSNSLWACAKLGYRD-SDLLLPLAEAAAALAPDMNAQGLANSLWALEVLGGTGPAFRAVLETLYGAALRRLRTPMGAEAFKPQ
NCC1      343 ALANSLYALGKLREDQQRRG-LGWDPTS-SPHLK-ALASAVASRLRAARGHGFTAQGVSNSLWACAKLGYRD-SALLLPLAEAAAALAPDMNAQDLANSLWALEALGCTRPEFRAVLKALCGAALRQLRTPKEAAAFIPR
NCL24     349 ALANALYALGKLREDQQQRG-SGWDPTS-SPHLK-ALASAVASRLRAARGHGFNAQDVSNSLWACAKLGYRD-SDLLLPLAEAAAALAREMNAQDVANSLWALEVLGCSGPAFRAVLEALCGAALRQLRTPKEAEAFKLQ
NCL22     340 ALANSLYALGKLREDTQQRG-SGWDPTS-SPHLK-ALASAVASRLRAARGHGFTAQGASNSLWACAKLEYRD-SDLLLPLAEAAAALAPDMNAQELSNSLWAWEALGCTGPAFRAVLETLYGAALRQLRTPKEAEAFKPQ
NCL25     336 DLANALYALGKLREDQQRRG-SGWDPTS-SPHLMNALTGAVASRLRAVSGHGFNAQDVSNSLWACAKLGYRD-SDLLLPLAEAAAALAREMNAQDVANSLWALEVLGCSGPAFRAVLEALCGAALRQLRTPKEAEAFKLQ
NCL17     342 ALANSLYALGKLREDQQRRG-SGWDPTS-SPHLK-ALASAVASRLRAARGHGFEPQHVSNSFWACAKLGYRD-SDLLRPLAEAAAALAPDMNAQGLANSLWALEVLGCTGSAFRAGLEALYGAALRRLRAPKDAAAFIPQ
NCL21     336 ALANALYALGKLREDQQRRG-SGWDPTS-SPHLK-ALASAVASRLRAARGHGFNAQDVSNTLWACAKLGYRD-SDLLRPLAEAAAALAPDMNAQGLANSLWALEVLGCTGSAFRAGLEAMYGAALRRLRTPKEAEAFKPQ
NCL19     337 DLANALYALGKLREDQQRRG-SGWDPTS-SPHLK-ALASAVASRLRAARGHGFNAQDVSNSLWACAKLGYRD-SDLLRPLAEAAAALAPDMNAQGLANSLWALEVLGCTGPAFRAGLETLYGAALRQLRTPKEAEAFKPQ
NCL20     337 ALANALYALGKLREDQQQRG-SGWDPTS-SPHLN-ALASAVASRLRAASGHGFEPQHVSNSLWACAKLGYRD-SDLLRPLAEAAAALAPDMNAQGLANSLWALEVLGCTGPAFRAVLETLCGAALRQLRTPKEAEAFKPQ
NCL15     120 NLANALYALGKLREQQQQRG-SGWDPTS-SPHLN-ALASAVVSRLHAAEGHGFSAQELSNSLWACAKLGYRD-SDLLLPLAEAAAALACGMRAQDLANNLWALEVLGCTGPAFRAVLETLCGAALRRLRTPKEAAAFIPQ
NCL35      93 ELSNALYALGKLREDHQQRG-SGWDPTS-SPYLN-ALAGAVASRLRAAEGHGFKPQELSNSLRACAKLGYRD-SALLLLSEAAAALAPDMSAQDLANSLWALDALGSTGPEFRAVLETLCGAALRQLRTPKEAEAFKPQ
NCL14     355 ELSNALYALGKLREEQQQRG-SGWDPTS-SPHLN-ALTGAVASRLRAAEGHVFKPQELSNSLWACAKLGYRD-SDLLLPLAEAAAALAPDMNAQEISNSLWALGALGCTGPAFRAVLETLCGAALRRLRVPEKSEAFKPQ
NCL1      341 NLANALYALGKLREQQQQRG-SGWDPTS-SPHLD-ALAGAVASRLRAAGGHGFTAQGVSNSLWACAKLGYRD-SDLLLPLAEAAAALALDMNAQDVAINLWALEVLGCTGSAFRAVLETLYGAALRQLRTPKEAEAFKPQ
NCL16     300 NLANALYALGKLWEEQQQRG-SGWDPTS-SPHLI-ALTGAVASRLRATEGHGFKPQHVSNSLWACAKLEYRN-SDLLLPLAEAAAA*
NCL30     341 NLANALYALGKLREDQQQRG-SGWDPTS-SPHLI-ALTGAVASRLRAARGHGFTAQNVSNSLWACAKLGYRN-SAILLPLAEAAAALAPDMNAQELANSLWALEALGCTGPAFRAVLETLYGAALRRLRAPKEAEAFKLQ
NCL34     342 NLANALYALGKLREDQQQRG-SDWDPTS-SPHLI-ALAGAVASRLRAAEVRGFTAQGVSNSLWACAKLGYGD-SDLLRPLAEATAALAPDMNAQGLVNSLWALEVLGCSGREFRAVLEALCGAALRQLQTPKETAAFKPQ
NCL33     340 NLANALYALGKLREDQQRRG-SGWDPTS-SPHLI-ALANAVASRWRGALGHGFKPQELSNTLWACAKLGYRD-SYLLLPLAEATAALARDMNAQDVANNLWALEVLGCTGPAFRAVLETLYGAALRRLRTPTEAAVFIPQ
NCL28     350 NLANALYALGKLREDQQQRG-SGWDPTS-SPHLI-ALAGAVASRLRAARGHGFEPQHVSNSLWACAKLGYRD-SDLLLPLAEATAALARDMNAQGLVNSLWALEVLGCTGPAFRAVLEALCGVALRQLRTPKGAELFKPQ
NCL29     345 NLANALYALGKLREDQQQRG-SGWDPTSSSPHLN-ALTGAVASRLRAARGHGFNAQDVSNSLWACAKLGYRD-SDLLLPLAEATAALARDMNAQGLVNSLWALEVLGCSGPAFRAVLEALCGAALRQLRTPKGAELFKPQ
NCL27     330 NLANALYALGKLREDQQQRG-SGWDPTS-SPHLI-ALAGAVASRLRAARGHGFNAQGVSNTLWACAKLGYRD-SDLLRPLAEATAALVLDMNAQELSNSLWALEVLGYTGPAFRAVLEAMCGAALRRLRTPKGAEAFKPQ
NCL31     336 NLANALYALGKLREDQQRRG-SGWDPTS-SPHLI-ALANAVASRLRAAEGHGFKPQELSNTLWACAKLGYRD-SYLLLPLAEATAALARDMNAQDVANNLWALEVLGCTGPAFRAVLETLYGAALRRLQTPKETAAFKPQ
NCL32     300 ELSNALYALGKLREDQQQRG-SGWDPTS-SPHLI-ALAGAVASRLRGALGHGFEPQHVSNTLWACAKLGYRD-SYLLLPLAEATAALARDMNAQDVANNLWALEVLGCTGPEFRAVLEALCGAALRRFQTPKETAAFKPQ
NCL26     244 DLCLVLYALGKLREDWQHTG-EGWQP-----CVR-LLAEAMKGRLAAAVGHGFNAQDVSNALWACAKLGYAD-ADLLQRLAEAGAAVAKTMIPQDLSNILWALKALGCTGPAYQPAVGALCEEALRRLRTPKEAEAFRPQ
NCL3      261 DLSLALYALGKLREDWQQNG-EGWD-----QSLG-KLADAIKARLTAAVGHGFNAQDVSNALWACAKLGLAD-AELLQRLAQAGAAAAGNMTPQALSNSLWALEALGCTGPAYRSAIAALCGEAFRRLRTPELAAAFAPQ
NCC2      257 DLSLALYALGKLREGWQQNG-EGWD-----QSLG-KLTDAIKTRLTAAVGHGFNAQDVSNSLWACAKLGLAD-ADLLQRLAEAGAAVAGDMTPQALSNSLWALEALGCTGPAYRAAVQVLCGEALRRLRTPKLAEAFKPQ
NCL5      258 DLSLALYALGKLREDWQRQGGAAWDPAE-SDSVG-ALADEIKGRLTAAVGHGFNAQDVSNSLWACAKLGLAD-AELLQRLAEAGAAAAGNMTPQALSNSLWALEALGCTGPAYRSAVQVLCGEALRRLRTPELAEAFNSQ
NCL6      121 VLSNAMYALGKLKQDQQQRGGSGWDPTS-SPRLQ-ILAAAVAHRLRVAVGHGFTAQGISNSLCACAKLGYRITAELLLPLAEAIASFSVSMQPHELATSLWALKALGCKGPDYRAAVEALCGEALQRLRTPKLAAAFTPQ
NCL13     249 ELSNVMYALGKLKQDQQQR-GSGWDPTS-SPHLQ-TLASAVASRLGAAVGHGFKPQELSNILWACAKLGYAD-PGLLLPLAEAAATLSVSMKVQELANSLWAMEALGCSGLEYRSAVEALCGEVLQRLRIPELAAAFTPQ
NCL10     252 SLTNVMYALGKIKQDQEQR-GSGWDPTS-SPHLQ-TLAAAVASRLGVAVGHGFTAQDVSNILWACAKLGYPE-PELLLPLAEAAAALSLSMETQHLSNTLWAFAALGCMGPEYRSAVEALCGEALRQLRAPKDAEAFKPQ
NCL7      251 NLAIVMYALGKLKQDQQQR-GSGWDPTS-SPHLQ-TLAAAVASRLRTVAGHGFKPQDVSN--------------------------------------SLWALKALGCTGPEYRSAVEALCGEALQQLRKPELAAAFIPQ
NCL8      148 NLAIVMYALGKLKQDQQQR-GSGWDPTS-SPHLQ-TLAAAVASRLRTVAGHGFKPQDVSN--------------------------------------SLWALKALGCTGPEYRSAVEALCGEALQRLRKPELAAAFIPQ
NCL9      251 NLAIVMYALGKLKQDQQQR-GSGWDPTS-SPHLQ-TLAAAVASRLRTVAGHGFKPQDVSN--------------------------------------SLWALKALGCTGPEYRSAVEALCGEALQQLRKPELAAAFIPQ
NCL11     261 ELSNAMYGTSKLRE-----------------------------------------------------------------------------------------------ALEYRSAVEALCGEALQQLRTPELAAAFKPQ
NCL12     182 ALSNAMYGTSKLRE-----------------------------------------------------------------------------------------------AFEYQSAVKALCGETLRRLRIPELAAAFTPQ
consensus 421  LanalYalgKlredqqqrg sgwdpts sphl  ala avasrlraa ghgf  q vsnslwacaklgyrd selllplaeaaaala  mnaq l nslwale lgctgpafra le lcgaalr lrtpk a afkpq
```

Supplemental Fig. 2A (continued)

```
                  OPR_E...                                                                                               ...OPR_E
               -----------                                                                                             ----------------------| |-----------
               ↓6                                                                                                                                ↓3 ↓6
NCL2      227 ELTNILMALQGLP-------------------------------------------------------------------------------------------QIGRQQSELLAAAVAAVDLQRDFAGYNSQDLSDSAW
NCL36     477 HLSNILLALEGL----------------------------------------------QLGGKQSELLAAAVAAECVRRLRTLKEAEAFIPQHLSNILLALEGLQLRGKQSELLASAVATEGVRRGFARFESQALSNSAW
NCL37     477 ELSNILLALEGL----------------------------------------------QLGGTQSELLAAAVAAECVRRLRTLKEAQAFKPQGLSNILLALEGLQLRGKQSELLASAVAAEGMRRAFAGFKPQDFSNAAW
NCL38     477 ELSNILLALEGLQLGGKQSELLAAAVVAECVRRLRTPKEAEAFKPQELSNILLALEGLQLGGKQSELLAAAVAAECVRRLRTLKEAEAFIPQELSNILLALEGLQLRGKQSELLASALAAEGMRRAFAGFKPQDFSNSAW
NCL23     450 GLSNILLALEGL--------------------------------------------------------------------------------------------QLRGKQSELLTAAVAAEGVRRGFAGFKPQELSNSAW
NCC1      479 HLSNILLALEGL--------------------------------------------------------------------------------------------QLRGKQSELLASAVAAEGVRRGFAGFNPQDLSNSAW
NCL24     485 GLSNILLALEGL--------------------------------------------------------------------------------------------QLRGMQAELLAAAVAAEGVRRGFAGFEPQHLSNSAW
NCL22     476 GLSNILLALEGL--------------------------------------------------------------------------------------------QLRGKQSELLASAVAAEGVRRGFAGFEPQHLSNSAW
NCL25     473 GLSNILLALEGL--------------------------------------------------------------------------------------------QLRGMQAELLAAAVAAEGMRRGFAGFKPQELSNSAW
NCL17     478 HLSNILLALEGL----------------------------------------------QLGGKQSELLAAAVAAECVRRLRAPKEAEAFIPQHLSNILLALQGLQLRGKPSELLADAVAAEGVQRGFAGFKPQELSNSAW
NCL21     472 DLSNILLALEGLQLGGKQSELLAAAVAAECVRRLRAPKEAEAFKPQELSNILLALEGLQLGGKQSELLAAAVAAECVRRLRAPKEAEAFKPQELSNILLALEGLQLGGKQSERLAAAVAAEGVRRGFAGFKPQELSNSAW
NCL19     473 HLSNILLALEGL--------------------------------------------------------------------------------------------QLGGKQSERLAAAVAVEAMRRGFAGCNPQELSNSAW
NCL20     473 HLSNILLALEGL--------------------------------------------------------------------------------------------QLGGKQSERLAAAVAVEAMRRGFAGCNPQELSNSAW
NCL15     256 HLSNILLALEGL----------------------------------------------QLRGMQAELLTAAVAAECVRRLGTPKEATAFIPQELSNILLALEGLKLRGKQSELLAAAVAAEGVQRGFAGFNPQCLSNSAW
NCL35     229 GLSNILLALEGL--------------------------------------------------------------------------------------------QLGGQQSERLAAAVAAEGVQRGFAGFNPQALSNSAW
NCL14     491 HLSNILLAMEGL----------------------------------------------QLRGMQAELLAAAVAAECVRRLRTPMGAEAFKPQELSNILLALEGLQLRGKQSELLAAAVAAEGVRRGFAGFNPQCLSNSAW
NCL1      477 DLSNILLALEGL--------------------------------------------------------------------------------------------QLRGKQSELLASAVAAACVRRGFAGFEPQHLSNSAW
NCL16
NCL30     477 GLSNILLALEGL--------------------------------------------------------------------------------------------KLRGKQSRLLAAAVAAECVRRGFAGFKPQHLSNSAW
NCL34     478 GLSNILLALEGL----------------------------------------------QLGGTQSERLAAAVAAECDRRLRARKEAEVFKPQELSNILLALEGLQLGGKQSERLAAAVAAEGVRRGFAGFEPQHLSNTAW
NCL33     476 HLSNILLALEGL--------------------------------------------------------------------------------------------QLGGKQSELLAAAVAAEGVRRNFAGFNPQALSNSAW
NCL28     486 ELSNILLALEGL--------------------------------------------------------------------------------------------QLGGKQSERLAAAVAAEGVWRGFAGFTPQDLSNSAW
NCL29     482 ELSNILLALEGL--------------------------------------------------------------------------------------------QLGSKQSERLAAAVAAEGVWRGFAGFTPQELSNSAW
NCL27     466 ELSNILLALEGL--------------------------------------------------------------------------------------------QLGGKQSELLAAAVAAEAMRRGFAGFNPQHLSNSAW
NCL31     472 GLSNILLALEGL--------------------------------------------------------------------------------------------QLGGTQSEQLAAAVAVEAMRRGFAGCNPQELSNSAW
NCL32     436 GLSNILLALEGL--------------------------------------------------------------------------------------------QLGGTQSEQLAAAVAVEAMRRGFAGCNPQELSNSAW
NCL26     376 HLSNILLALEGL--------------------------------------------------------------------------------------------QLGGKQAELL-AAVATEDLRRGFSGYNDQDLTNSAW
NCL3      393 HLSNILLALEGL--------------------------------------------------------------------------------------------QLGREKAQLV-AAVAADGVRRGFAGFKPQELSNSAW
NCC2      389 DLSNILLALEGL--------------------------------------------------------------------------------------------QLGSEQAQLV-SAVAAEDVRRGFTRYNSQDISNSTW
NCL5      395 HLSNILLALEGL--------------------------------------------------------------------------------------------QLGSDQAQLTV-AVAAEDVRRSLAGYVAQDISNSAW
NCL6      259 DLSTILLALEGL--------------------------------------------------------------------------------------------QLGGKQAELV-AAVAAEDVRRGFAGYGAQNVGNSAW
NCL13     385 HLSNILLALEGL--------------------------------------------------------------------------------------------QLGGKQAELV-AAVAAEDVRRGFDGYVAQDLSNSAW
NCL10     388 HLSNILLALEGL--------------------------------------------------------------------------------------------QLGGEQAELVV-AVAAEDVRRGFAGYVAQDLSNSAW
NCL7      350 QLSNILLALEGL--------------------------------------------------------------------------------------------QLCSEQAELV-VAVAEEDMRRGFDGYIAQDLSNSAW
NCL8      247 QLSNILLALEGL--------------------------------------------------------------------------------------------QLCSEQAELVV-AVAAEDVRRGFDGYIAQDLSNSAW
NCL9      350 QLSNILLALEGL--------------------------------------------------------------------------------------------QLCSEQAELVV-AVAEEDMRRGFDGYIAQDLSNSAW
NCL11     306 ELSNVLLALEGL--------------------------------------------------------------------------------------------QLGGGQAGLVV-AVAAEDVRRGFTGYVPQDLNNSAW
NCL12     227 ELSNILLALEGL--------------------------------------------------------------------------------------------QLGGKQAELV-AAVAAEDVRRGFEGYVPQDLSNSGR
consensus 561  lsnillalegl                                                                                             ql g qse laaavaaegvrrgfagy pqdlsnsaw
```

Supplemental Fig. 2A (continued)

```
                 OPR_F                                                OPR_G                                 OPR_H                                   OPR_I
               -------------------------|             |----------------------------------|--------------------------------------|    |---------------------
                                                        ↓3 ↓6                                 ↓3 ↓6                                    ↓3 ↓6
NCL2      276 ALAKMGYGQGTTPLSSKQSQWYAAAVAAAQR---PGVMASAKPQAWANLLYALALVRYQPSPTLLEAGAASVMES---GTAHSCAITLWALAVLQLRHAGVESAVCDRLGELLWLHPASVPVQHLSSSLWALAVLAGG-G
NCL36     571 ALAKMGYGAGSPSEAMEQRQWYAAAVAAAQR---PGVMASAKPQAWANLLYALALVRHQPPPVLLDAGAAAAMQR---GNGQECANTLWALAVMRLRHAGLEAAVCGRLGELLRLGPESLAVQELCNSLWALAVLAGG-G
NCL37     571 ALAKMGYGAGSPSQAMEQRQWYAAAVAAAQR---PGVMASAKPQAWANLLYALALVRHQPPPALLDAGAAAAMQR---GSAQECANTLWALAVMQLRHAGLEAAVCGRLGELLRQDPESLVVQNLCNSLWALAVLAGG-G
NCL38     617 ALAKMGYGAGSPSEAMEQRQWYAAAVAAAQR---PGVMGSAKPQAWANLLYALALVRHQPPPALLDAGAAAAMQR---GSAQACANTLWALAVLQLRHAGLEAAVCGRLGELLRQDPELLAAQALCNSLWALAVLAGGGG
NCL23     498 ALAKMGYGAVATPQATEQSHWYAAAVAAAQR---PGVMAGATPQNWANLLYALALLRHQPPPALLDAGAAAAMQR---GNAQECANALWALAVLQLRHAGLEAAVCGRLGELLQREPESLIAQELCNSLWALAVLAGG-D
NCC1      527 ALAKMGYGAGATPQAMEQSHWYAAAVAAAQR---PGVMAGAKPQDWANLLYALALVRHQPPPTLLDAGAAAAMQG---GNAQECSNTLWALALLQLRHAGLEAAVCGRLGELLRQGPESLIAQELCNSLWALAVLAGG-G
NCL24     533 ALAKMGYGAGATPQATEHSHWYAAAVAAAQR---PGVMAGATPQNWANLLYALALVRHQPPPALLDAGAAAAMQR---GNAQHCANALWALAVLQLRHAGLEAAVCGRLGELLRQGPESLAEQALCNSLWALAVLSGG-G
NCL22     524 ALAKMGYGVGATPQAMEQSPWYAAAVAAAQR---PGVMAGAKPQDWANLLYALALVRHQPPPVLLDASAAAAMQR---GSAQHCANSLWALAVLQLRHVGLEAAVCGRLGELLRQGPESLAEQDLCNGLWALVALAGG-G
NCL25     521 ALAKMGYGAGATPQAMKQCPWYAAAAAAAQR---PGVMAGATSQNWANLLYALALVRHQPPPALLDAGAAAAMQR---GNAQDCANTLWALAVLQLRHAGLEAAVCGRLGELLRQEPESLAEQALCNSLWALAVLAGG-G
NCL17     572 ALAKMGYGAGATPQAIEQSHWYAAAVAAAQL---PGVMAGAKPQAWANLLYALALVRHQPPPTLLDAGAAAAMQG---GNAQACANTLWALAVLQLCHAGLEAAVCGRLGELLRLGPESLVAQALCNSLWALAALAGG-G
NCL21     612 ALAKMGYGAGATPQAMEQSHWYAAAVAAAER---PGVMAGAKPQAWANLLYALALVRHQPPPTLLDAGAAAAMQR---GTAQHCANTLWALAVLQLRHAGLEAAVCGRLGELLRLGPESLVEQALCNSLWALAALAGG-G
NCL19     521 ALAKMGYGAGATPKAMEQCPWYAAAVAAAQR---PGVMASAKPQHWANLLYALALVRHQPPPALLDAGAAAAMQR---GKAQECANTLWALAVLQLRNAGLEAAVCGRLGELLQREPESLVAQALCNSLWALAVLAGG-G
NCL20     521 ALAKMGYGAGATPKAMEQCPWYAAAVAAAQR---PGVMASAKPQAWANLLYALALVRHQPPPALLDAGAAAAMQR---GKAQECANTLWALAVLQLRNAGLEAAVCGRLGELLQREPESLVAQALCNSLWALAVLAGG-G
NCL15     350 ALAKMGYGASATSQATQQSHWYAAAAAAAQR---PGVMAGAMPQNWANLLYALALVRHQPPPALLDAGAAAVMQR---GNAQDCANTLWALAVLQLRHAGLEAAVCGRLGELLRLEPESLVAQELCNSLWALAVLAGG-G
NCL35     277 ALAKMGYGAGATPQATEQSHWYAAAVAAAER---PRVMAGAKPQEWANLLYALALVRHQPPPALLDAGAVVAMQR---GNAQECANTLWALAVLQLRHAGLEAAVCGRLGELLRLGPESLVAQNLCNSVWALAVLAGG-G
NCL14     585 ALAKMGYGAGATSQATEQSQWYAAAAAAAER---PGVMAGAKPQEWANLLYALASVRHQPPPALLDAGAAAVMQR---GNAQECANTLWALAVLQLRHAGLEAAVCGRLGELLRLGPESLVAQELCNSLWALAVLAGG-G
NCL1      525 ALAKMGYGAGATSQATQQSHWYAAAAAAAQR---PGVMAGAKPQGWANLLYALALVRHQPPPALLDAGAAAVMQR---GNAQECANTLWALAVLQLRHAGLEAAVCGRLGELLRLEPESLIAQELCNSLWALAVLAGG-V
NCL16
NCL30     525 ALAKMGYGAGATLHAAEQRHWYAAAVAAAQR---PGVMAGAKPQAWANLLYALALVRHQPPPALLDAGAAAAMQR---GNAQACANTLWALAVLQLRHAGLEAAVCGRLGELLRQGPESLVAQDLCNSLWAVAVLAGG-G
NCL34     572 ALAKMGYGAGATPQAIEQRPWYAATVAAAQR---PGVMAGAKPQDWANLLYALALVRHQPPPALLDAGAAAAMQR---GNAQHCANTLWAQAVLQLRHAGLEAAVCGRLGELLRLDPESLVAQALCNSLWALAVLAGG-G
NCL33     524 ALAKMGYGAGATPKAMEQSPWYAAAVAAAQR---PGVMASAKPQYWANLLYALALVRHQPPPALLDAGAAAAMQR---GNVQECANTLWALAVLQLRHAGLEAAVCGRLGELLRQGPESLTAQELCNSLWALAVLAGG-G
NCL28     534 ALAKMGYGAGATSQAIEQCPWYAAAVAAAQR---PGVMASAKPQAWANLLYALALVRHQPPPALLDAGAAVAMHR---GNAQECANTLWALAVLQLRHAGLEAAVCGRLGELLRQGPESLIAQELCNSLWALAVLAGG-G
NCL29     530 ALAKMGYGADATSQAIEQSPWYAAVAAAAQR---PGVMAGAKPQAWANLLYALALVRHQPPPALLDAGAVAAMQR---GNAQECANTLWALAVLQLRHAGLEAAVCGRLGELLRQGPESLIAQELCNSLWALAVLAGG-G
NCL27     514 ALAKMGYGAGATSQAMKQSPWYAAVTAAAQR---PGVMAGAKPQAWANLLYALALVRHQPPPALLDAGAAAAMQR---GDAQACANTLWALAVLHLCHAGLEAAVCGRLGELLRQGPESLIAQELCNSLWALAVLAGG-D
NCL31     520 ALAKMGYGAGATSQAMKQSPWYAAVAAAAQR---PGVMAGAKPQAWANLLYALALVRHQPPPALLDAGAVAAMQR---GNAQDCANTLWALAVLQLRHAGLEAAVCGRLGVLLRLEPESLAAQNLCNSLWALAVLAGG-G
NCL32     484 ALAKMGYGAGATSQAMKQSPWYAAVAAAAQR---PGVMAGAKPQAWANLLYALALVRHQPPPALLDAGAVAAMQR---GNAQDCANTLWALAVLQLRHAGLEAAVCGRLGELLRLGPESLAEQNLCNSLWALAVLAGG-G
NCL26     423 ALAKMEFGVGPE-APEEQRKWISAAVDAAKRRGRDGVMAGATPQAWSNLLYALALVRHQPPPALLDAGAAAAIL--KSGNGQDCASVLYALAVLRLRHGELEAAVCGRLGALLKRDKQAVNHQCIATSLWGIAVP----G
NCL3      440 ALAKMGFGAGPE-APEEQRQWVSAAVDAAMR---PGTMATATPQAWANLLYALALMRHQPPPALLQNAGAAAAML-SSGSGQNCANTLYALAVLQLRHAGLEAAVCGRLRELLRGDRQALNNQEVSNTLWAMAVL----G
NCC2      436 ALAKMGFGVGPE-APAEQRQWVTAALDAAMR---PGTMATATPQNWSNLLYALSVMRHQPPPVLLDAGAAAAMRANSVNGPQDCANTLYALAVLQLRHAGLEAAVCGRLGELQQEDLESLTEQGLANSLWAVAVF----G
NCL5      442 ALAKMGFGVGPE-APEEQRQWVTAALDAAMR---PGTMATATPQAWANLLYALALMRHQPPPVLLDAGAA-AMRANGVNG-QECANTLYALAVLQLRHAGLEAAVCGRLGELLWGGKQALNHQDIANSLWAMAVF----G
NCL6      306 ALAKMGLGVGPE-APTEQRQWVSAAVDAAMR---PGVMARATPQAWANLLYALALTRHQPPPGLLDAGAGESMQR---GNGQDCANTLWALAVLQLRHAGLEAAVVGRLGELLQHDSQALNQQDISSSLWAMAVL----G
NCL13     432 ALAKMGFGVGPE-APVEQRQWVSAAVDAAMR---PGVMARATPQAWANLLYALSMIRHQPSPALLDAGAAAAMHR---GNAQDCANTLWALAVLQLRHAGLEAAVVGRLGELLQRGGEALTEQGISNSLWAMAAF----G
NCL10     435 ALAKIGFGVGPE-APAEQRQWASAAVVAAMR---PGVMARATPQAWANLLYALALMRHQPPPELLDAGAAAAAMRASSAGGQNCANTLWALAVLQLRHAGLEAAVCGRLGELLQRDRQAVNNQNISNSLWAMAVI----G
NCL7      397 ALAKMGFGAGPE-APAEQRQWVSAAVDAATQ---SGAMARATSQAWSNLLYALALMRHQPPAELLDSGAAVAAMRASSAGGQNCANTLYALAVLQLRHAGLEAAVCGRLGELLQHDSDTVLHQHVCNSLWAMAVF----G
NCL8      294 ALAKMGFGAGPE-APAEQRQWVSAAVDAATQ---SGAMARATSQAWSNLLYALALMRHQPPAELLDSGAAVAAMRASSAGGQNCANTLYALAVLQLRHAGLEAAVCGRLGELLQHDSDTVLHQHVCNSLWAMAVF----G
NCL9      397 ALAKMGFGAGPE-APAEQRQWVSAAVDAATQ---SGAMARATSQAWSNLLYALALMRHQPPAELLDSGAAVAAMRASSAGGQNCANTLYALAVLQLRHAGLEAAVCGRLGELLQHDSDTVLHQHVCNSLWAMAVF----G
NCL11     353 ALAKMGFGVGPE-APAEQRQWVSAAVDAAMR---PGVMTGATPQEWANLLYALALMRHQPPPALLDAGAAA-AMRASSADGQDCANTLYALAVLQLRHAGVEAAVCGRLGELLQRGGKALTEQGISNSLWAMAVL----G
NCL12     274 VALY*
consensus 701 alakmgygag    a eq  wyaaavaaa r   pgvma a pqawanllyalalvrhqpppalldagaaaamqr   g aq cantlwalavlqlrhagleaavcgrlgellr  pesl  q lcnslwalavlagg g
```

Supplemental Fig. 2A (continued)

```
                                                              OPR_J
               ------------------|          |-----------------------------------------|                                                              |-----
                                               ↓3 ↓6
NCL2      409 GPAGLAAAPLARALAEEAV-RRRKDL---TKGGLAQLWQARQELGNEV-EALTRGPDVQAAMEAAVAATQ-ATGSNPSRTQEQAAKALLSLTQK------GLLPIVSVWTETAVEGMLGRVDIVTDWSFGRMVAVEVDGP
NCL36     704 GPASPAATALAPALAREAV-RRREGP---QNGGLCQLWQARQELGGEVAEALARSPDLNAAMEAAVAAEW-ATKSNTSSTQEQVAEALKRLLQK------GRLPIVSVETEVVVEGVLGRADIVAGWSDGRRVAIEVDGP
NCL37     704 GPASPAIAALAPVLAREAA-RRREGF---QIEGLCQLWQAQQELGGKVAEALARSPDLNAAMEAAVAAER-ATKSTTSSTQEQVAEALKRLLQK------GCLPIGFVQTEVVVEGVLGRADIVAGWSDGWRVAIEVDGP
NCL38     751 GPASPAATALASALAREAV-RRREGF---QIEELCQLWQVQQELGGKVAEALARSPDLNAAMEAAVAAER-ATKSTTSSTQEQVAEALKRLLQK------GCLPIGFVQTEVVVEGVLGRADIVAGWSDGRRVAIEVDGP
NCL23     631 GPASPAASAVAPALAREAA-RRREGF---QTEALRQLWQAKHELGGEVTEALARSPDLLAAMEAAVAAER-ATGSTTSSLQEQVAEALQRLLRK------ARLPVVSVRTEVVVEGVLGRVDIVADWSDGRRVAIEVDGP
NCC1      660 GPASPAAAAMAPALAREAA-RRREGF---QTEGLLQLWQAQHELGGEVAEALARSPDLLAAIEAAVAAVQ-ATGSTTSSLQEQVAEALQRLLQK------GRLPIVSVRTEVVVEGVLGRVDIVADWSDGRRVAIEVDGP
NCL24     666 VPASPAAAALAPALTREAA-RRREGF---PTGALSQLWQAQHELGGEVTEALARSPDLLAAMEAAVSAKR-ATGSTTSSLQDQVAEALQRLLQK------GRLPIVSVRMEVVVEGVLGRVDIVADWSDGRRVAIEVDGP
NCL22     657 GPASPAAAALAPALAREAA-RRCEEL---QAEALSQLWQAQHELGGEVTEALARSPDLLAAMEAAVAAKR-ATGSNTSRLQEQVAEALQRLLQK------GRLPIVSVRTEVVVEGVLGRVDVVADWSDGRRVAIEVDGP
NCL25     654 GPASPAAAALAPALAREAA-RRHEEF---QTEALSQLWQAQQELGGEVTAVLARSPDLLAAMEAAVAAER-VTGSNTSRLQEQVAEALQRLLQK------GRLPIVSVRTEVVVEGVLGRVDIVADWSDGRRVAIEVDGP
NCL17     705 VPASPAAAALAPALAREAV-RRREEL---QTEALSQLWQAQHELGGEVAEALARSPDLLAAMEAAVADER-ATESNTSRLQEQVAETLQRLLQK------GRLPIVSVRTEVVVEGVLGRVDVVADWSDGRRVAIEVDGP
NCL21     745 VPASPAAAALAPALAREAV-HRIRDL---AGDDFRQLWQAQHELGGEVAEALARSPDLLAAMEAAVAAQR-STTSTTSSTQKQVAEALQRLLQK------GRLPIVSVRTEVVVEGVLGRVDIVADWSDGRRVAIEVDGP
NCL19     654 VPASPAASALAPALVREAV-RRREEF---QTEELRQLWQAQHELGGEVAEALARSPDLLAAMEAAVSAKR-ATGSTTSSTQEQVAEALQRLLQK------GRLPIVSVRMEVVVEGVLGLVDMVADWSDGRRVAIEVDGP
NCL20     654 VPASPAASALAPALVREAV-RRREEF---QTEELRQLWQAQHELGGEVAEALARSPDLLAAMEAAVSAKR-ATGSTTSSTQEQVAEALQRLLQK------GRLPIVSVRMEVVVEGVLGRVDIVADWSDGRRVAIEVDGP
NCL15     483 GPASPAAATLASALAREAA-RRREEF---QTEALSQLWQAQHALGGEVTEALARSPDLLAAMEAAVAAER-ATTSNTSSTQKQVAEALQRLLQK------DLLPIVSVQTEAAVEGVLGRVDIVADWSDGRRVAIEVDGP
NCL35     410 VPASPAAAALAPALAREAV-RRREQFVDMSAEHLRQLWQAQHELGGEMAEALARSPDLNAAMEAAVAAER-ATGSNTSSTQKQVAKALQRLLQK------DLLPIVSLQTEPVVEGVLGRVDIVADWSDGRRVAIEVDGP
NCL14     718 GPASPAAAALAPALALEAV-RRREEL---NGDGLRQLWQARQELGGEVAEALARSPDLLAAMEAAVAAER-ATTSTTSRLQEQVAEALQRLLQK------DLLPIVSVQTEVVVEGVLGRVDIVAGWRDGRSVAIEVDGP
NCL1      658 GPASPAAAALAPALAYEAV-RRREKL---NGDDFSQLWQARQELGGEVAEALARSPDLLAAMEAAVAAER-ATTSTTSSTQEQVAEALQRLLQK------DLLPIVSVQTEVVVEGVLGRVDIVADWRDGRRVAIEVDGP
NCL16
NCL30     658 GPASPAASALAPALAREAV-RRRQEF---QTEGLLQLWQARQELGGEVAEALARSPDLLAAMEAAVAAER-ATGLNTSRLQEQVAEALQRLLQK------GRLPIVSVRTEVVVEGVLGRVDIVADWSDGRRVAIEVDGP
NCL34     705 GPASPAASALAPALAREAV-RRREGF---QTEGLLQLWQAQQELGCEVAEALARSPDLLAAMEVAVVAER-ATGSTTSRLQKQVAEALQRLLQK------GRLPIVSVQTEVVVEGVLGRVDIVADWSDGRRVAIEVDGP
NCL33     657 GPASPAAAALAPALAREAA-RRRQGF---QTEELRQLWQARQELGGEVAEALARSPDLLATMEAAVAAKR-ATESNTSRLQEQVAEALQRLLQK------GRLPIVLMQTEVVVEGVLGRVDIVADWSDGRRVAIEVDGP
NCL28     667 GPASPAAATLASALAREAA-RRREEF---QTEALSQLWQARQELGGEVAEALARSPDLNAAMETAVAAEW-ATESNTSRLQEQVAEALQRLLQK------GRLPIVSVRTEVVVEGVLGRADMVADWSDGRRVAVEVDGP
NCL29     663 VPASPAAAALALALAYEAV-RRREEV---QTEELFQLWQAQQELGGEVAEALARSPDLLAAMEAAVATQR-ATVSTTSSTQEQVAEALQRLLQK------GRLPIGSVQTEVVVEGVLGRVDIVADWSDGRRVAIEVDGP
NCL27     647 GPASPAAAALAPALAREAV-RRREER---NGEYLCQLWQARQELGGEVAEALARSPDLLAAIEAAVAAKR-ATESTTSRTQEHVAEALRRLLQK------GRLPIVLVQTEVVAEGVLGRADIVADWSDGRRVAIEVDGP
NCL31     653 GPATPAAAALAPALAREAV-RWREEL---NGDGLRQLWQARQELGGEVAEALARSPDLLAAIEAAVAAKR-ATESNTSRTQEQVAEALRRLLQK------GRLPIVLVQTEVVAEGVLGRVDIVADWSDGRRLAIEVDGP
NCL32     617 GPATPAAAALAPALAREAV-RWREEL---NGDGLRQLWQARQQLGGEVAEALARSPDLLAAMEAAVAAER-ATGLNTSRTQEQVAEALRRLLQK------GRLPIVLVQTEVVAEGVLGRVDIVADWSDGRRLAIEVDGP
NCL26     556 GPLSAGMQQLAVEVARDAA-GRWEEFT--ATEDLAQMWQAQQELGGEVAAALGSNGKLQAAMDAVVAARR-DNSKPLPDDHKQLLAALRRLEQHGGGATAGGLALESVQTGVVMSGVLAPLDAVVRLSDGRQVALELVGA
NCL3      571 GTSSPGMQQLAIQLARNAA-GRWEGLI---HEDLSQLWQAQQELGGEVAETLCGISSLQAAMDKSVETYR-QDTKRLSETHKQLLAALRRLEQHQGREAAGGFAVVSVQAGVVAPGVLAAVDAVVRLNDGRQVAVELAGA
NCC2      568 ETGSPAMQQLAMQLARDAA-IRWEEFA---DEGLTQLWQAQQALGGEVAAALRGNRSLQAAMDKAVATYR-EDTKHLPDDQKQLLAALRRLEQHGRET-AGGLAVQSVQTGVVAQGVLTPVDAVMGLVDGRQVAVEMLGP
NCL5      572 ETCSPDMRRLAIQLARDAA-SRWDDFT---IEELTQLWQAQQELGGEVAAALRDSSSLQATMDAVVAARR-EETKPLHANQKQLIGTLRRLEQHGRETAAGGLSIQSVQADAVVPGVLAPMAAVVGLSDGRQVAVEWIGL
NCL6      435 GTSFPGMQQLAIELARDAA-IRWEGFT--AVAGLTQLWQAQRELGGEVAAALGGSPGLQAAMAAAVAAERAADVKPPPDNQKQLLAALRRLEQQGLITTTEGPSIQSLQTGVILPGILAPVAAVVGLSDGQQAAVELVGL
NCL13     561 NTCSPGMQQLAVQLARNAAGRRWKGFT---DDGLRQLWQAQQELGGEVAAALGSSPGLQAAMDKSVETYR-QDTKRLSETHKQLLAALRRLE---GLATAGGLAVQSVQTGVVAPGVLTPVDVLVGLSDGRQVAVEWVTV
NCL10     567 GTYSPDMQQLSIQLAREAV-SRWEGFT---TENLNQLWQAQQELGGEVAAALGSSPGLQAAMAAAVTAER-ENAKPTSDIQKQVVAALRRLEQHQGREAAGGLAVVSVQTGVVAQGALAPVDAVVRLSDGQQAAVELLGP
NCL7      529 GTYSPDMKQLSIQLAREAV-SRWEGLT---TEHLKQLWQAQQELGGEVAAALGSGSGLQAAMAAAVTTER-EDAKPTSDTQKQVVAALRRLEQ-QGLKAAGGLAVQSLQTGVVAPGVLAPVDAVVRLSDGQQAAVEFLGP
NCL8      426 GTYSPDMKQLSIQLAREAV-SRWEGLT---TEHLKQLWQAQQELGGEVAAALGSGSGLQAAMAAAVTTER-EDAKPTSDTQKQVVAALRRLEQ-QGLKAAGGLAVQSLQTGVVAPGVLAPVDAVVRLSDGQQAAVEFLGP
NCL9      529 GTYSPDMKQLSIQLAREAV-SRWEGLT---TEHLKQLWQAQQELGGEVAAALGSGSGLQAAMAAAVTTER-EDAKPTSDTQKQVVAALRRLEQ-QGLKAAGGLAVQSLQTGVVAPGVLAPVDAVVRLSDGQQAAVEFLGP
NCL11     484 GTSFPGMQQLAVQLARDAA-GRWEGFS---GEGHRQLWQAQQELGGEVAAALCSSPGLQAAMAAAVTAER-EDAKPTSDIQKQVVAALRRLEQHQGREAAGGLAVRSVQTGVFAPGVLAPVDAVVGLSDGRQVAVELVGP
NCL12
consensus 841 gpaspaa  la alarea  rrre       e l qlwqaqqelggevaealarspdl aameaavaa r at s ts  q qvaeal rllqk      g lpivsvqtevvvegvlgrvdiva wsdgrrvaievdgp
```

Supplemental Fig. 2A (continued)

```
                              RAP Domain
               ----------------------------------------------------------|
NCL2      537 IHFLRGRKG---NVSAVDGSTALRNRQLQRAFGKG-NVLYVPYWHWNGLKTPAKQEAYLLRRLQQ
NCL36     833 DHFLTNRKD---DPSAVDGSTALRNRQLRRALGEG-RLLCVPYWEWYGLRTPSAQEAYLLQQLQDVLFGASSGAA-----AGEGIA--AAPRRQQQRTTQPQPDVAGSTTTTT--A------STSDTAGSSQKRRVL-FV
NCL37     833 DHFLTNRKD---DPSAVGGSTALRNRQLRRALGAG-GLVCVPYWEWYGLRTPSAQEAYLLQQLQDVLFGASSGAA-----AGEGFP--AAPRRQQQRTTQPQPDVAGSTTTTTATA------STRNTAGSSQKRRVL-VV
NCL38     880 DHFLANRKD---DPSAVGGSTALRNRQLRRALGEG-GLVCVPYWEWYGLRTPSAQEAYLLQQLQDVLFGASSGAA-----AGEGIA--AAPRRQQQRTTQPQADIAGSTTTST--V------STSDTTGSSQKRRVH-IV
NCL23     760 DHFLTNRKD---DPSAVVGSTALRNRQLRRALGEG-GLLCVPYWEWNRRKTSSAQEAYLLQQLQDLLSGASSGAS-----AGEGSA--AARRRKEQRTTQPQPYTARSPTSASTTA---------IAAGSSHKRRIL-VV
NCC1      789 DHFLTNRKD---DPSAVNGSTALRNRQLQRALGEG-GLVCVPFWEWYGLRTPSAQEAYLLQRLQDLLSGASSGAA----G-GGGSA--AAPRHQQQRTTRLQPDIAGSSTTS-----------TTSTANSSQKRRVL-VV
NCL24     795 DHFLTHRKD---DPSAVIGSTALRNRQLRRAFGEG-GLLCVPYWEWDRQKTSSTQEAYLLQQLQDLLSGASSGTA----AAGEGSA--AARRRQQQRTTRLQPDTANSSTTTSTTT------ATTTTAGSSQKRRIL-VV
NCL22     786 DHFLIDRKD---DPSAVVGSTALRNRQLRRALGEG-GLLCVPYWEWDRQKTSSAQEAYLLQRLQDLLSGASSGAS-----AGYGTA--AARRRKEQRTTRLQPDTANSPTSASTAA------STSTTASSSQKRRIL-VV
NCL25     783 DHFLTNRKD---DPSAVDGSTALRNRQLRRAFGEG-GLLCVPYWEWDRQKTSSAQEAYLLQQLQDLLSGASSGAS-----AGEGSA--AAPRHQQQRTAQPQPDTANSSTT-----------STTSTASSSHKRRVL-IV
NCL17     834 DHFLTNRKD---DPSAVDGSTALRNRQLRRALGEG-GLLCVPYWEWYERQAPSAQKAYLLQQLQDLLSGASSGAS-----AGEGSA--AARRRKEQRTTQPQPYTAGSNTSASTTASIT------TTASSSQKRRIL-VV
NCL21     874 DHFLTNRKD---DPSAVDGSTALRNRQLRRALGEG-GLLCVPYWEWYERQAPSAQEAYLLQQLQDLLSGASSGAS-----AGEGSA--AARRRKEQRTTQPQPYTARSPTSASTTASITTTATTTTTASSSQKRRIL-VV
NCL19     783 DHFLTHRKD---DPSAVIGSTALRNRQLRRAFGEG-GLLCVPYWEWYGLRTPSTQEAYLLQQLQDLLSGASSGTA-----AGKGSA--AAPRHQQQRTAQPQPDTANSS----------TTATTTTTASSSQKRRIL-VV
NCL20     783 DHFLTHRKD---DPSAVIGSTALRNRQLRRAFGEG-GLLCVPYWEWYGLRTPSTQEAYLLQQLQDLLSGASS*
NCL15     612 DHFLTNRKD---DPSAVDGSTVLRNRQLRWAFGEG-GLLCVPYWEWDRRKTYSAQQAYLLQQLQDLLSGASSGAA----SAGEGSA--ATPRRKQQRTTPPQPDTAGRSSSTTNTT----STSASTTAGSSQKKRVLVVV
NCL35     542 DHFLTTRKD---DPYAVIGSTALRNRQLRRAFGRG-GLLCVPYWEWDRRKTPSAQEAYMLQQLQDLLSGASSGA-----AAGEGSA--ATPRRKQQRTIPPQPDTAGSSSITTATT--------STTASSSQKRRVLV-V
NCL14     847 DHFLTNRKN---DPSAAIGSTALRNRQLRRAFGEG-GLLCVPYWEWYGLKTSSAQEAYLLQQLQDLLSGASSGAV----AAGEGSA--ATPRRKQERTTRPQPNTAGRTSTS------------T-TAGSSQNRRVLVVV
NCL1      787 DHFLTNRKD---DPSAVNGSTVLRNRQLRRAFGQG-GLLCVPYWEWDRRKTSSAQEAYMLQQLQDLLSGASSGA-----AAGEGSA--AAPRRPQQRTTPPQPDTAGSSTTT------------STTASSSQKRRVLVVV
NCL16
NCL30     787 DHFLTNRKD---DPSAVDGSTALRNRQLRRAFGEG-GLLCVPYWDWCGRKTPSFQEAYLLQQLQDLLSGAPSGAA----AGGEGSA--AAPRRKQQRTTQPQPDTAGTTTT------------ATTTASSSQKRRIL-VV
NCL34     834 AHFPTNRKD---DPSAVIGSTALRNRQLRRAFGEG-GLVCVPYWEWYGLRTPTAQEAYLLQRLQDLLSGASSGAA----AGGEGIA--AAPHCKQQRTTRPQPDTAGSSTTTASTT-----TITSTTAASSEKRRIL-VV
NCL33     786 DHFLTNRKD---NPSAVDGSTALRNRQLRRTLGEG-GLVCAPYWEWYSLRTPSAQEAYLLRRLQDLLSGASSGGA----AAGEGIA--AAPRRQQQRTTRPQPDTAGTTTS------------TSTTAVSSQERRIL-VV
NCL28     796 DHFLTNRKD---DPSAVDGSTALRNRQLRRAFGEG-RLLYVPYWEWDIRKTPTAKEAYLLQRLQDLLSGASSDTA-----AGKGNA--AAPRRQQQRTTQPQPDTAGTTTT------------ASTTASSSQKRRIL-VV
NCL29     792 DHFLTNRKD---NPSAVDGSTALRNRQLRRTLGER-GLVCAPYWEWYSLRTPSAQEAYLLRRLQDLLSGAPSGAA----AGGEGSA--AAPRRKQQRTTRPQPDTAGTTTT------------ATTTASSSQKRRIL-VV
NCL27     776 DHFLTNRKD---DPSAVVGSTALRNRQLRRAFGEG-GLLCVPYWEWYGRKTSSAQEAYLLQQLQDLLSGASSGAA----A-GEGSA--AARRRKQQRTTRPQLDTAGITTST---T-----TSTTTTAGSSQKRRIL-VV
NCL31     782 DHFLTNRKD---DPSAVIGSTALRNRQLRRAFGEG-GLLCVPYWEWDDRQTPSFQEAYLLQRLQDLLSGAPSGAA----AGGEGSA--AAPRRKQQRTTRLQPDTAGTSTA------------ASTTASSSQTRRIL-VV
NCL32     746 DHFLTNRKD---DPSAVIGSTALRNRQLRRAFGEG-GLVCVPYWEWDERQTPSFQEAYLLQRLQDLLSGAPSGAA----AGGEGSA--AAPRRKQQRTTQPQPDTAGTTTT------------ATTTAVSSQTRRIL-VV
NCL26     692 VRFLSNRR---RDPAAVNGSTAMRHRHVRREYGEG-GVLMVPYWEWESLTSPEEQEAYLLRRLQALLLATAQRAV-------------GTPTRTTAG--D--AASPT-------------------SKSSSCGSPQLL--A
NCL3      706 VSFLANLRQ--RDTAAVNGGAAMRHRQLRRAFGEG-GVLLVPYWEWDRLQTAEEQEAYLLRRLQKVVVMAAEPQAAEAATAGAGRAAAGAPVRTTGA--TASVVSPSSNNSNSS--------RGSPGGGSPGRSRPL--A
NCC2      702 KRFIYNRKQ--DDPTAVNGGTGMRNRQLRRAFSEG-GVLLVPHWEWEGLKSPEEQEAYLLLRLQEVAAA-AETQAAAAAAQGAGKTSARTSAAGAAA--APAPARPGSSGSSSS--------RGGRGGGSPGSPQLL--T
NCL5      707 KRVLRSRQR--EGAILLEGGIGMRIRQLQRAFREG-GVLVVPHWEWEGRRSPEEQETYLLRRLQEVVAA-AETQAAAAAAKGAGKAPAGTPAAGAAA--APVPARPGSSSSS-------------PGSGSPGSPQLL--A
NCL6      572 KRYLYNRKQPQ-DPADVDGGTAMRIRQLRRAFGEGGGVLVVPYWEWAGLKSPEEQEAYLLHRLQEPMAPPPEVVG-------NGGI--RASE--ETR--SAAP---AASSPAAA--------TSGDGAAGNK*
NCL13     694 VRFLSNRKQ---DPSAVNGSTVLRHRQLRRAFGEG-GVLLVPYWEWDGLQTAEEQEAYLLRRLQQPAVAVETVT--------AGVV--GGAA--AVI--TAAPHQPATTTTTTT--------KGGGGSGSNPQQQLL--V
NCL10     702 ARFLSNLKR---DPTAVDGGTAMRIRQLRRAFGEGGGVLVLPYWEWAGLKSPEEQEAYLLRRLQQPAVAVETVA--------VGAV--GGAA--AVI--TAAPQQPATTTT-TS--------KGGGGSGGNPQQQLL--V
NCL7      663 KRYLLTQTQQQDPSAAVEGGTAIRIRQLRRAFKEG-GVLLVPYWEWDFL-TPSEQEAYLLRRLQQPAVAVETVT--------AGAV--GGAA--AVT--TAAPHQPATTTTNSN--------KGGGGSGGNPQQQLL--V
NCL8      560 KRYLLTQTQQQDPSAAVEGGTAIRIRQLRRAFKEG-GVLLVPYWEWDFL-TPSEQEAYLLRRLQQPAVAVETVT--------AGAV--GGAA--AVT--TAAPHQPATTTTNSN--------KGGGGSGGNPQQQLL--A
NCL9      663 KRYLLTQTQQQDPSAAVEGGTAIRIRQLRRAFKEG-GVLLVPYWEWDFL-TPSEQEAYLLRRLQQPAVAVETVT--------AGAV--GGAA--AVT--TAAPHQPATTTTNSN--------KGGGGSGGNPQQQLL--A
NCL11     619 TRFLSDQTCS--GRSAVDGRTVLRIRQLRRAFGEGGGVLVVPYWEWDALTN-AEEKAYLLRRLQQPAVARAVTV---------GAE--GGAA--AVT---------AAATTTTA--------KGGGGSGGNPQQQLL--V
NCL12
consensus 981 dhfltnrkd   dpsav gstalrnrqlrrafgeg gllcvpywew   ktps qeayll  lqdvl ga s         g g a  aa rr   rt   qp  agsttt                a ss  rrvl  v
```

Supplemental Fig. 2A (continued)

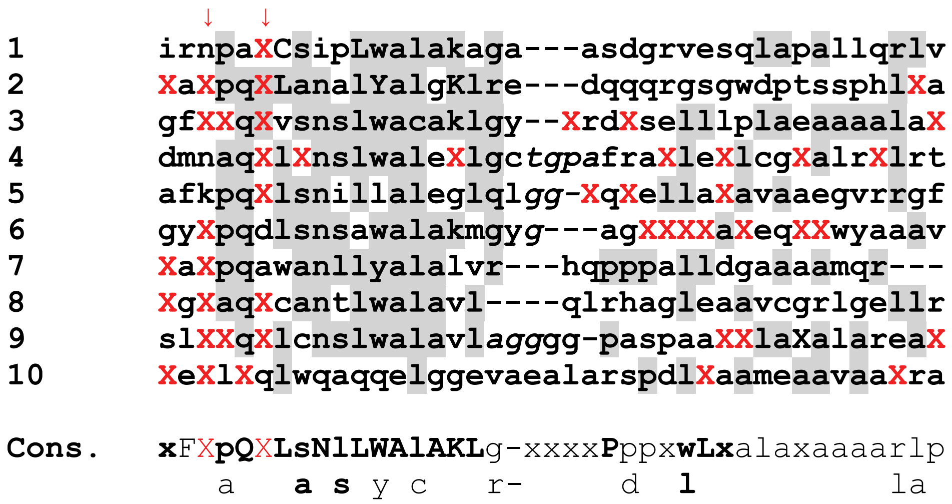
```
NCL2
NCL36      953 RRKAPKQEEAAGSGG-AGGASGATATA-----AAGAELVAPAGQRQAVEAQRLPNSGAGGGAGEGGGDGTPLAPQPPPPQ-RQR--------VSAPRARRNSSQSRSRSSQAGSPLPPP-PSPVKAPPPEQVQ--AAAPP
NCL37      955 RRKTPKQEA-PGGGG-AGEASGATAAA-----AAGAELAAPAGQRQAEEAQRLPSLGASGGAGEGGGDGTPLAPQPPPPQ-RQR--------VSAPRARYSSSQSRSRSSQAGSPSPAP-P-PVEAPPPEQVQ--AAAPP
NCL38     1000 RRKTPKQEA-AGGGG-AGGASGATAAA-----AAGAELAAPAGQRQAEEAQRLPSSGAGGGAGEGGGDGTPLAPQPPPPQ-RQR--------VSAPRARYSSSQSRSRSSQAGSPSPAP-P--VKAPPPEQVQAQAAAPP
NCL23      879 RGKAPK-QEAVDGGG-AGGASNATAAAA-TVAVAGAEVVAPAGQQQAEESHQLPNIGAGGGAGEGGEDGTPLAPQPPPPQ-QQH--------VSAPRARRSSSQSRSRSSQAGSPSAPP-PPPVEAPPPEQMQ--AAAPP
NCC1       906 RRKALK-QEAAGGGG-AGGASNATAAAA-TVAVAGAESVAPAWQQPVEKAQRLPSSGAGGGAGEGGEDGASLAPQPPPPQ-QQH--------VSAPRARRSSSQTQSRSRSS*
NCL24      918 RRKAPKQQEAAGGGG-AGGGSDATAAAV-AAGVAGAESVAPAWQQPVEKAQRLPSSGAGGGAGEGGEDGASLAPQPPTPQ-QQR--------VSAPRARRSSSQSQSRSSQAGLPAPAPAPPSVKAPPLEQTR--QLNPG
NCL22      908 RKKAPM-QEAAGGGG-AGGASGATATAA-TI-PAEAKVVAPALQQQAEESHQLPNIGAGGGAGEGGGDGTPLAPQPPPPQ-QQR--------VSAPRARRSSSQSRSRSSQAGSPSPPPPPPPVQAPPPEQMQ--VAAPP
NCL25      900 RRTARK-QEAAGGGG-AGEASGATAAAR-TV-RAEAEVAAPARQQSAEKAQRLPSSGAGGGAGEGGGNGTPLAPQPPPTQ-QQR--------VSAPRARRSSSQSRSRSSQSGSP--PP-APPVEAPPPEQTQ--AAAQS
NCL17      956 RRKAPK-QEAVGGGG-AGGASSATAAAA-TI-PAEAKVVAPALQQQAEESHQLPSSGAAGGAGEGGGDGTPLALQPPPPQ-QQH--------VSAPRARRSSSQSRSRSSQAGSPSAPP-PPPFEAPPPEQVQ--AAAPP
NCL21     1002 RRKAPK-QEAVGGGG-AGGASSATAAAA-TI-PAEAKVVAPALQQQAEESHQLPSSGASGGAGEGGGDGASLALQPPPPQ-QQH--------VSAPRARRSSSQSRSRSSQAGSPSAPP-PPPFEAPPPEQVQ--AAAPP
NCL19      901 RRKAPK-QEAAGGGG-AGEASGATAAAR-TV-PAEAEVAAPARQQSAEKAQRLPSSGAGGGAGEGGGNGTPLAPQPPPPQ-QQR--------VSAPRARRSSSQSRSRSSQAGSPS--P-APPVEAPPPEQVQ--AAAPP
NCL20
NCL15      738 RRKAPKQEAVGGGGG-AGGASDATVDAA--TVPAEAEVAAPARQQQAEESQRLPISGAGGGAGEDGGDGAPLAPQPPPPQQRPR--------VSAPRARRSSSQSHSSSSSSQA--GSPPPPSVEAPPPEQVQ--AAAPP
NCL35      662 RRTAPKQEAVGTSGG-AGGASDAAAATATAAAAGAKSVAPAWQQQAEEAQRLPSS---GGAGEGGGDGAPLAPQPPPPQ-QPR--------VSAPPTRSSSRQS*
NCL14      964 RRKARKQQEVACGGG-AGEASDATVAAA--TVPAEAEVVAPAGQQQTETVQRLPNSGAGGGAGEGGGDRAPLAPQPPPPQ-QPR--------VSAPRARRSSRQSPSRS---IQ--AGSPPPSVEAPPPEQMQ--AAAPP
NCL1       904 RRTA-PKQEAAGGGG-AGEASDATVAAR--TVPAEAEVVAPAGQQQTETVQRLPSSGAGGGAGEGGGDGAPLAPQPPPPQ-QPR--------VSAPRTRRSSRQSQSRG---VQ--AGLPAPSVEAPPPEQMQ--AAAQP
NCL16
NCL30      904 RRKAPKQDLQHDATCDAPSASGAGPAE------AVAEVA---------------TSG---GP-RPATAGADVASAPRRQQRQRQL--QQQPTRPEPEAVAAAV-GRRRRPAAK---EPRPQQQQTRPPPQNTRKRAASAN
NCL34      958 RRKARKQDLQHNATFDTPSASGAGPAE------AVAEVA---------------TSG---GL-RPAAAGADVAPAPRQQQRQRQ---HQQPTRPEPEAVAATV-SRRRRPAAN---EPRPQQ-QQRPPPQNTRKRAAASN
NCL33      903 RRTARKQDLQHDATCDTPSASGAGPAE------AVPEVA---------------TSG---GL-RPAAAGADVAPAPRQQQRQRQQ*
NCL28      912 RRKAPKQDLQHDATCDAPSASGAGPAE------AVAEVA---------------TSG---GL-RPAAAGADVASAPRQQQRQRQ---LQQPSRPEPEAVAVAV-GSCSRPAAN---EPRPQQ-QTRPPPRNTRKRAASAN
NCL29      909 RRTAPKQDLQHDATCDAPSASGAGPAE------AVAEVA---------------TSG---GL-RPAAAGADVASAPGRQQRQRQ--RQQQPSMPEPEAMAAAV-GRRRRPAAN---EPRPQQQQTRPPPQNTRKRAASAN
NCL27      896 RRTARKQDLQHDATCDAPSASGAGPAE------AVAEVA---------------TSG---GP-RPATAGADVAPAPRQQQR------QQQPSRPEPEAVAAAV-GRRRRPAAK---EPRPQQQQERPPPRNTRKRAAFAN
NCL31      899 RRKAPKQDLQHDATFDAPSASGAGPAE------AVAEVT---------------ASG---GS-RPAAAGADVASAPRRQQRQRQLQQQQQPSRPEPEAVAAAV-GRRRRPAAK---EPRPQQQQERPPPQNTRKRAASAN
NCL32      863 RRKAPKQDLQHDATCDAPSASGAGPAE------AVAEVA---------------TSG---GP-RPAAAGADVAWAPGQQQQQRQ-RQLQQPSRPEPEAVAAAV-GRRRRPAAK---DPQPQQQQTRPPPQNTRKRAAASN
NCL26      791 R-PPRR-VP*
NCL3       831 R-SPRR-V*
NCC2       826 R-PPRR-RV*
NCL5       826 R-PPRR-RV*
NCL6
NCL13      806 R-PARR-PEGGGGG*
NCL10      814 R-PARR-PEGGGGGGGSGRGGR*
NCL7       777 R-PARR-PEGGGGRRRVGGS--KRTVGLRS*
NCL8       674 R-PARR-PEGGDGGGGGGGGSKRTVVGLRS*
NCL9       777 R-PARR-PEGGDGGGGGGGGSKRTVVGLRS*
NCL11      724 R-PARR-PEGGGGGGGGSGRGGR*
NCL12
consensus      r  a k      g   a  as a          a                          g      g   la  p   q
```

Supplemental Fig. 2A (continued)

```
NCL2
NCL36    1075 RRAAGEKRTERKAAAAPAAGGGG--GGGGGSDPGAVVVPAAVSEKQVAVVAGEEGSTAAAVKGGD--GGGSGGDGGLPAGSRRRRAVGTGRRAASSSLAAATEQMRLEQ*
NCL37    1075 RRAAGGKRTERKAAAAPAADGGG--GGGGGTDPGAVVVPAAVSEKQVAVVAEEEGSTAAAVKGGD--GGGSGGGGGLPAGSRRRRVVGTGRRAASSSLAAASEQMRLEQ*
NCL38    1121 RRAAGGKQTV---AVDPAAGGGG--GGGGGTDPGAVVVPAAVSEKQVAVVAGEEDSSAAAVRGGD--GGGSGGGGGLPAGSRRRRAVGTGRRAASSSLAAASEQMRLEQ*
NCL23    1004 RRAAGGKRTERKAAAATAAGGGG--GGGGGSDPGAVVVPAAVSEKRLSVVAGEADSSAAAVECGGGH----GSGGGLLAGSRRKRAVGTGRRAAQKPSSSE*
NCC1
NCL24    1045 GPRVGSGLSGKQ---------------RRPRQQAAVAAAAAVTPGRWWFRPRSLRSDS----------------RWWPGRRTRRQQL*
NCL22    1033 RRAAGGKWTERKAAAAPAAGGGG--GGGGGSDPGAVLVPAAVSEKR*
NCL25    1022 RRAAGGKRTERKAGVAPAAGDG----GGGGSDPGAVVVPAAVSEKQVVVVAGEADSLAAAVECGGG-GDGSGNGGGLLAGSRRGQAVGTGRRAAQKPSSSE*
NCL17    1080 RRAAGGKRTERKAAAAPAAGGGGGGGGGGGSDSGAVVVPAAVSEKRLSVVAGEADSSAAAVECGGGGGDGSGSGGGLPAGSRRGQAVGTGRRAAQKPCSIE*
NCL21    1126 RRAAGGKRTERKAAAAPAAGGGG-GGGGGGSDSGAVVVPAAVSEKR
NCL19    1023 RRAAGGKRTERKAAAAPAAGGGG---GGGGSDPGAVVVPAAVSEKRLSVVAGEADSSAAAVECGGGG---DGSGGGLPAGSRRGQAVGTGRRAAQKPCSIE*
NCL20
NCL15     863 RRAAGGKRTERKAAAAPAAGGGG----GGGSDPGAVVVPAAVSENQVAVVAGEEGSTVAAVAGGG-GGGGSGGGGGLPAGSRRRRAGGTGRRAAQKSSSSV*
NCL35
NCL14    1085 RRTAGGKRTVRKAAAAPAAGG-------GGSDPEAVVVPAAVSEKQAAVVAGEEGSTLADVEGGG-SGSGRGGGSRRR-----RAV------GPAGVRPRQA*
NCL1     1024 RRAAGGKRTERKAAAAPAAGGGG--GGGGGSGPGAVVVPATVSENQVAVVAGEEGSTAADVEGGG-GGGGSGGAGACP-----RAAGGGGRRGPAGVQPRSVPGTSE*
NCL16
NCL30    1013 A-------HAAAAAAAAAA*
NCL34    1065 A-------DAAAAAA*
NCL33
NCL28    1019 A-------HAAAAAAAAAA*
NCL29    1018 A-------HAAAAAAAAAA*
NCL27    1001 A-------HAAAAAAAAAA*
NCL31    1010 A-------HAAAAAA*
NCL32     973 A-------HAAAAAA*
NCL26
NCL3
NCC2
NCL5
NCL6
NCL13
NCL10
NCL7
NCL8
NCL9
NCL11
NCL12
```

Supplemental Fig. 2A (continued)

B)



```
                  ↓  ↓
1        irnpaXCsipLwalakaga---asdgrvesqlapallqrlv
2        XaXpqXLanalYalgKlre---dqqqrgsgwdptssphlXa
3        gfXXqXvsnslwacaklgy--XrdXselllplaeaaaalaX
4        dmnaqXlXnslwaleXlgctgpafraXleXlcgXalrXlrt
5        afkpqXlsnillaleglqlgg-XqXellaXavaaegvrrgf
6        gyXpqdlsnsawalakmgyg---agXXXXaXeqXXwyaaav
7        XaXpqawanllyalalvr---hqpppalldgaaaamqr---
8        XgXaqXcantlwalavl----qlrhagleaavcgrlgellr
9        slXXqXlcnslwalavlagggg-paspaaXXlaXalareaX
10       XeXlXqlwqaqqelggevaealarspdlXaameaavaaXra

Cons.    xFXpQXLsNlLWAlAKLg-xxxxPppxwLxalaxaaaarlp
            a   a s y c   r-     d  l          la
```

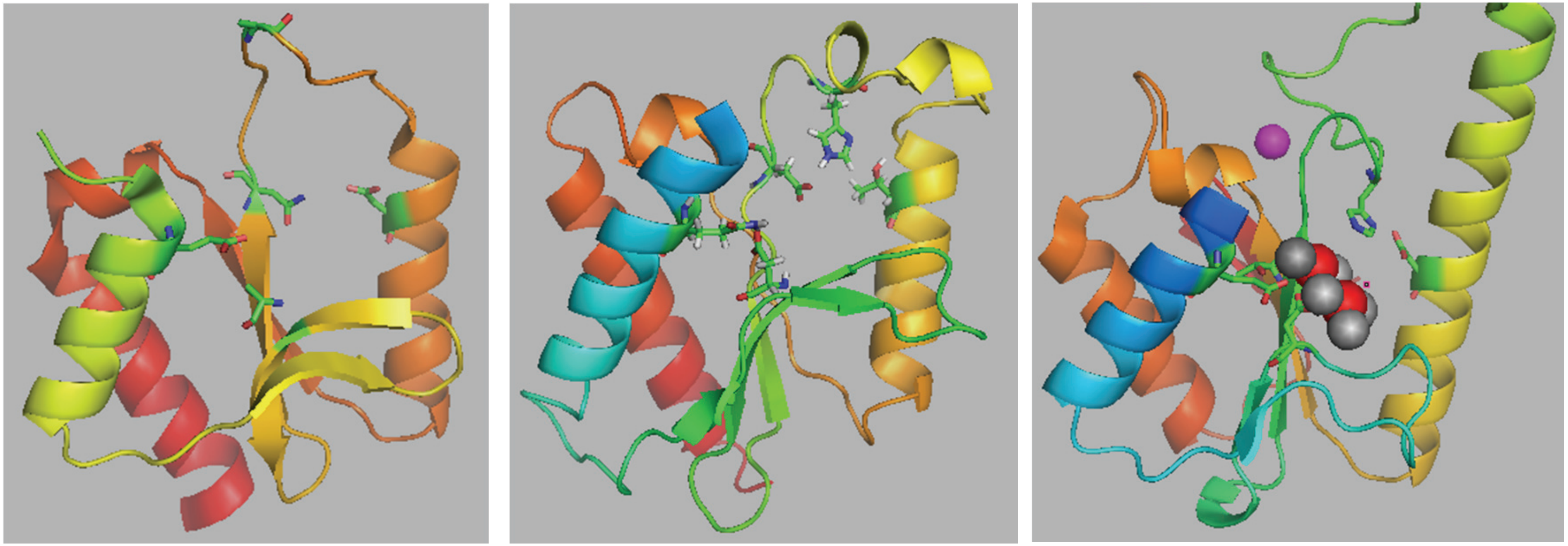
Supplemental Figure 2: Alignment of NCCL proteins.

A) Alignment was done with the clustal□ software, and manually edited to improve alignment. It was edited with BoxShade (rtf old option: http://www.ch.embnet.org/software/BOX_form.html). Residues conserved in more than 60 % of the sequences (>22 out of 36) are written in red, conservative substitutions are indicated by blue letters, while amino acids at variable positions or differing from the consensus are written in black. Residues written in grey are those that show similarity to other NCL proteins but are not translated because of premature Stop codons. The positions of the OPR repeats and of the RAP domain are shown above the sequence. Variations in the number of the OPR repeats mostly result from intra-protein repeats deletions (e.g. OPR_C in NCL7, NCL8 and NCL9; OPR_B and OPR_C in NCL11 and NCL12) or duplications (see the region of OPR_E in NCL14, 15, 17, 34, 36-38, with NCL38 showing two identical copies of the repeat AFKPQELSNILLALEGLQLGGKQSELLAAAVAAECVRRLR). Arrows point to positions 3 and 6 within the OPR repeats, which show a higher variability compared to the rest of the repeat, suggesting that these positions are under diversifying selection. Residues mutated in NCC1 and NCC2 are black-boxed and written in white. Red arrowheads point to the position of introns.

B) Consensus sequences for each OPR repeat, retrieved from panel A), were aligned. Residues in capital letters are found at that position in all NCL proteins, while positions less conserved (a same amino acid was not found in, at least, 23 out of 36 NCCL proteins) are indicated by a red X. Residues fitting the OPR consensus, shown below the alignment, are shaded in grey. Arrows point to position 3 and 6 that show the highest variability within the OPR repeats, written in red in the consensus.

The RAP domain may show endonucleolytic activity.

The structural modelling of the RAP domain of NCC1 and NCC2 by the I-TASSER software (Zhang, 2008) used four endonucleases as major templates. Two of them (3r3p, the Catalytic Domain of Homing Endonuclease I-Bth0305I (Taylor et al., 2011) and 3hrl, a putative endonuclease-like protein from *Neisseria gonorrhoeae*) are annotated as carrying a DUF559 domain, also found in many bacterial OPR-like proteins. The other two (1cw0, the very short patch repair (Vsr) endonuclease (Tsutakawa et al., 1999; Bunting et al., 2003) and 4oq2, the 5hmC specific restriction endonuclease PvuRTs1I (Kazrani et al., 2014)) are grouped with DUF559 in the cl00277 superfamily of the CDD database. This superfamily corresponds more or less to the PE..(D/E)xK family of endonucleases, named after the catalytic residues (Knizewski et al., 2007). The models produced for the RAP domains (Fig S4) indeed show a mostly antiparallel β-sheet sandwiched between two α-helices, similar to the αβββαβ topology of PE..(D/E)xK endonucleases. The catalytic residues of the endonucleases align reasonably well with residues that are well-conserved in an alignment of 640 RAP domains; Fig. S5). The RAP domain could have conserved the metal-dependent endonuclease function of its ancestors, with a specificity shifted towards RNA rather than DNA. Mechanistically, this is not unrealistic: the monomeric Vsr cleaves a single strand of the DNA, next to a T-G mismatch, while the others act as dimers to generate double stranded cuts. Interestingly, the sole OPR protein in *Arabidopsis*, required for the proper processing of *rrnS* precursor (Kleinknecht et al., 2014), also contains a RAP domain.



Structure of 3r3p

Model generated for NCC1 by I-TASSER

Structure of 1cw0 (protein + metals)

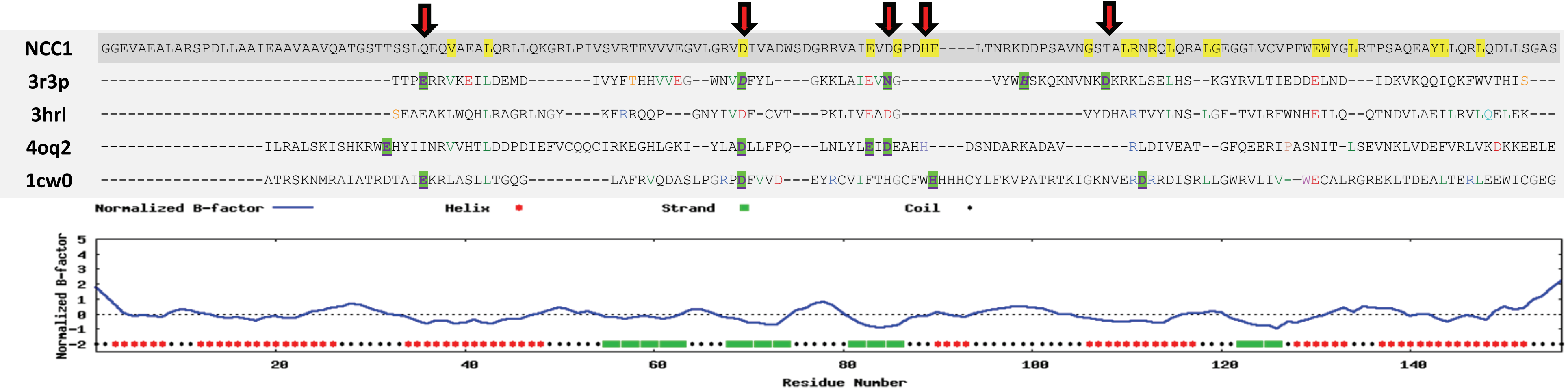
(res 1-31 removed)

Fig. S3: Comparison of NCC1 model with known structures of endonucleases

Residues involved in catalysis (experimentally defined for 3r3p and 1cw0, predicted for the RAP domain) are shown in stick representation, colored by element. In 1cw0, the Zn atom is shown in magenta, the Mg atoms in red and their bonding water oxygen atoms in grey

Similarity between the RAP domain and endonucleases.

Fig. S4: I-TASSER alignment used for threading of the NCC1 RAP domain.



```
NCC1  GGEVAEALARSPDLLAAIEAAVAAVQATGSTTSSLQEQVAEALQRLLQKGRLPIVSVRTEVVVEGVLGRVDIVADWSDGRRVAIEVDGPDHF----LTNRKDDPSAVNGSTALRNRQLQRALGEGGLVCVPFWEWYGLRTPSAQEAYLLQRLQDLLSGAS
3r3p  --------------------------------TTPERRVKEILDEMD-------IVYFTHHVVEG--WNVDFYL----GKKLAIEVNG----------VYWHSKQKNVNKDKRKLSELHS--KGYRVLTIEDDELND---IDKVKQQIQKFWVTHIS---
3hrl  --------------------------------SEAEAKLWQHLRAGRLNGY----KFRRQQP---GNYIVDF-CVT---PKLIVEADG--------------------VYDHARTVYLNS-LGF-TVLRFWNHEILQ--QTNDVLAEILRVLQELEK---
4oq2  ------------------ILRALSKISHKRWEHYIINRVVHTLDDPDIEFVCQQCIRKEGHLGKI--YLADLLFPQ---LNLYLEIDEAHH----DSNDARKADAV-------RLDIVEAT--GFQEERIPASNIT-LSEVNKLVDEFVRLVKDKKEELE
1cw0  ------------------ATRSKNMRAIATRDTAIEKRLASLLTGQG---------LAFRVQDASLPGRPDFVVD---EYRCVIFTHGCFWHHHHCYLFKVPATRTKIGKNVERDRRDISRLLGWRVLIV--WECALRGREKLTDEALTERLEEWICGEG
```

Residues highlighted in yellow correspond to those best conserved in an alignment of 640 RAP domains from Uniprot. Those highlighted in green have been identified as important for catalysis in the endonucleases (in 3r3p, italicized D196 and H213 have been mutated to A). In the template sequences, colored residues correspond to those also found in NCC1. Candidate active site residues in NCC1 are indicated by arrows. The normalized B-factor and predicted secondary structure are shown at the bottom.

3r3p: Homing Endonuclease I-Bth0305I Catalytic Domain (Taylor et al, 2011); 3hrl: putative endonuclease-like protein (ngo0050) from *Neisseria gonorrhoeae* (unpublished); 4oq2: 5hmC specific restriction endonuclease PvuRTs1I (Kazrani et al, 2014); 1cw0: Vsr endonuclease (Bunting et al, 2003; Tsutakawa et al, 1999)

Best identified stuctural analogs in PDB:

	PDB Hit	TM-score	RMSDa	IDENa	Cov.
NCC1	1cw0A	0.580	3.00	0.138	0.705

Ranking of proteins is based on TM-score of the structural alignment between the query structure and known structures in the PDB library. RMSDa is the RMSD between residues that are structurally aligned by TM-align. IDENa is the percentage sequence identity in the structurally aligned region. Cov. represents the coverage of the alignment by TM-align and is equal to the number of structurally aligned residues divided by length of the query protein.

Supplemental Table 1: Genetic independence of the *ncc2* mutation from *MCA1* and *TCA1* genes.

A	**Phenotype**	**Number of descendants**
	WT	24
	Leaky b_6f	22
	b_6f mutant	34

B	**Tetrad type**	**Number of tetrads**
	Parental Ditype (PD)	4
	Recombinant Ditype (RD)	2
	Tetratype (T)	5

A. Analysis of descendant phenotypes for the cross *ncc2* x *mca1*-2
B. Analysis of tetrad types for the cross *ncc2* x *tca1*-2

Supplemental Table 2: Markers designed to map the *ncc2* mutation on chromosome 15

marker		primers	position on chromosome 15	Ann. T. (°C)	elong. time (mn)	taille 137c (pbs)	taille S1-D2 (pbs)
ZYS3	F:	AGCCGCCACGTGTTTGTGGAGG	225,954..225,975	58	2	350	500
	R:	ACTGCCTTCTGGCTCGTATGCGGG	226,281..226,304				
343	F:	CGGCCAGCCCACTACCTGC	343,100..343,118	57	2	130	110
	R:	GCATATCTGCACATGCACATGTAC	343,213..343,236				
382	F:	CAGCTTCAGCCGCAGCAACAGCA	382,423..382,445	58	1	550	600
	R:	AAGGGACATGCATTCGCGCTCAGC	382,966..382,989				
425	F:	AGCAGGTGGGGTTTTCTAGGCGGCGTG	425,174..425,200	60	2	460	several smaller bands
	R:	GCGTACAGGTGAACAGCAGGGTGTGAT	425,622..425,648				
433	F:	CTCTGGCGCTGGCGGGGCATG	433,993..434,013	62	0.5	140	150
	R:	GCCCTCAAACTCGCGCAATGCCGA	434,104..434,127				
457	F:	AGGGTATGCCTTGATCGCACACATAC	457,538..457,563	60	0.5	120	90
	R:	CATGCCATGTCTGCTGTACACCC	457,629..457,651				
467	F:	CTTGCGTGTTTGCTGCCGACACCG	467,160..467,183	60	1	500	-
	R:	CCGAGGCGGTGCAGGTGGGTG	472,764.. 472784 [a] 476,779..476,799				
480	F:	ACCGAACCCTCCGGCTCCTCATCAACACC	480,488..480,516	60	0.5	250	-
	R:	CCCGGCCGAGCGCTTCGGCGACCTCTTA	480,706..480,733				
497	F:	TCGCCCTACCCCCACGTCCA	497,948..497967 [a] 501694..501,713	56	2	920	620
	R1:	TCACACCCATGCAATGACCTAG	CACW27613.fwd .30..288 [b]				
	R2:	CCCTAACTGAGCACATAATTCC	498,846..498,867				
509	F:	CACCAAATCCCGTTCGCACTCCTC	509,758..509,776	55	0.67	320	120
	R:	ATGCGGCGCCCAACCGGAG	509,758..509,776				
513	F:	AGTCGGGCCAACTGCCGGAAG	513,236..513,256	55	0.67	290	150
	R:	AAGCTAAGCACTGATCAAACATCCACATTAACATG	513,497..513,528				
531	F1:	ATAGGACCAGGCTTAGGGCCCT	531,550..531,571	55	0.67	390	220
	F2:	CGCGTACCGGCTGGAGGG	CACW4379.fwd 407..424[b]				
	R:	CTGCCACAATACCTGCCACAGCA	531,899..531,921				
576	F:	CGTTAAGGCGGCGGCTTACATCG	576,016..576,038	55	0.5	420	300
	R:	CGCGTCCAGGCGCTCATCCC	576,420..576,439				
623	F:	TACGCCCTCCAGCCCAAGTCCTG	623,359..623,381	58	0.5	300	-
	R:	TGGCGGCGGCTACGTAATGTAC	623,633..623,654				

1063	F: GCCACGTTGCAACCTCGCTTG	1,063,557..1,063,577	58	0.5	210	190
	R: ATGTGAGGGCGGCTGCGGG	1,063,557..1,063,577				
1114	R: TCTGCTGCAATGCTGATGCC	1,114,852..1,114,871	55	0.67	350	250
	F1: TATGTTGCATTTACGACGGTGAG	1,115,195..1,115,217				
	F2: TGGCAGGGACAGCTGTGC	CET1|178473|925063A1 315..298[b]				

Columns describe, from left to right, the name of the marker, the sequence (5' to 3') of the corresponding oligonucleotides, their position on chromosome 15 or on S1-D2 ESTs, the annealing temperatures and elongation times used for PCR amplification and the size of the amplicons in *C. reinhardtii 137c* and in S1D2.

- indicates that no PCR product could be amplified from S1D2 with the indicated pair of primers.

[a]: this oligonucleotide has multiple hits on the *Chlamydomonas* genome, but only that one gives rise to a detectable PCR amplification product under the experimental conditions used.

[b]: these oligonucleotides do not align along the *Chlamydomonas* genome but along EST from strain S1D2.

Supplemental Table 3: Oligonucleotide used in this study

Name	Sequence 5' to 3' [a]	*Restriction sites* [b]
AFRF_FW	CGC**CCATGG**CTATGTCTAACCAAGTATTTACTACT	***NcoI***
AFRF_RV	CGC**CTGCAG**TTAGAAGTTCATTTCTGCTAATTGAACT	***PstI***
atpA$_{\text{Fus}}$FW	CGC**AAGCTT**CAATGCGTACTCCAGAAGAACT	***Hind*III**
atpA$_{\text{Fus}}$RV1	CGC***CTGCAG***ACCATCACCTAATAATACCGCA	***PstI***
atpA$_{\text{Fus}}$RV2	GCT***CTGCAG***TTAAGCTGTCATACTACCTTCA	***Pst*I**
*atpA*Cod	CAATGCGTACTCCAGAAGAACT	
*ncc1*_FW	CGTTCATTCCGAGACACCTG	
*ncc1*_RV	TGCGGCTACGGCTTTGGGTT	
*ncc2*_FW	TATGACACCACAGGCACTCA	
*ncc2*_RV	CCAGACCTTGCTCAGTCAGT	
atpA$^{\text{M}}$FW	AACTCAGGTTTACGCCCAGCG**GTTAAC**GTAGGTATTTCAG	***Hpa*I**
atpA$^{\text{M}}$RV	CTGAAATACCTAC**GTTAAC**CGCTGGGCGTAAACCTGAGTT	***Hpa*I**
atpA$_{\text{Ext}}$FW	AGGTAGTATGACAGCTCTTC**CAATTG**TTGAAACACAAGAAGGT	***Mfe*I**
atpA$_{\text{Ext}}$RV	TATACCGAGTTCCTAC**TTAATTAA**AAATTAAGCAGCTTTAGCT	***Pac*I**
petA$^{\text{M}}$FW	ACAAAACTACGCTAACC~~CGCGAG~~AAGCAAACGGTCGTATTGTATGTGCA	~~*Pml*I~~
petA$^{\text{M}}$RV	TGCACATACAATACGACCGTTTGCTT~~CTCGCG~~GGTTAGCGTAGTTTTGT	~~*Pml*I~~
petA$_{\text{Ext}}$FW	TCCCCATTTTTATAA**AGATCT**TCCATGCATGAACT	***Bgl*II**
petA$_{\text{Ext}}$RV	TGGGATTTT**GTCTAC**AACAACTTCACCGTTTGCTTTT	***Acc*I**
*petD*5::T2_FW	TAG**ATTTAAAT**CCACGTGAGGCTAATGGTGTAATTCTGTCCCTTTTTAC	***Swa*I**, *Pml*I
petD$_{\text{Cod}}$::T2_RV	GGC**CTGCAG**CCATTAGCCTCACGTGGAATACACCTAGAAGTTTGTTTG	***Pst*I**, *Pml*I

[a]: bases written in red differ from the wild type *Chlamydomonas* sequence

[b]: Restriction sites introduced in the sequence of the oligonucleotide for cloning purpose (written in bold) or for RFLP analysis of the transformants (written in black). The crossed-out restriction site points to a restriction site, originally present in the wild type *petA* sequence but destroyed in the transformant.

Supplemental Table 4: *NCC1* and *NCC2* paralogous genes

E value (/NCC2)	Score (/NCC2)	gene model		
0.0	1643.3	NCC2	618.3	0.0
0.0	1102.8	Cre15.g640450.t1.2	584.3	0.0
0.0	930.2	Cre15.g640350.t1.2	584.3	0.0
0.0	818.5	Cre15.g638450.t1.2	568.2	0.0
0.0	723.4	Cre15.g639700.t1.1	676.4	0.0
0.0	722.6	Cre15.g637850.t1.1	657.9	0.0
0.0	692.2	Cre15.g639802.t1.1	609.4	0.0
0.0	690.3	Cre15.g639750.t1.1	622.1	0.0
0.0	687.6	TTT15.g639800.t1.1	608.6	0.0
0.0	666.8	Cre15.g639850.t1.2	632.1	0.0
0.0	595.9	Cre15.g638303.t1.1	1032.3	0.0
0.0	584.7	Cre15.g638300.t1.1	969.1	0.0
0.0	578.6	Cre19.g750697.t1.1	974.2	0.0
0.0	570.1	Cre15.g638150.t1.1	932.9	0.0
0.0	567.4	Cre15.g638100.t1.1	930.2	0.0
0.0	567.0	Cre15.g638956.t1.1	954.9	0.0
0.0	565.1	Cre19.g750497.t1.1	926.8	0.0
0.0	562.4	Cre04.g223150.t1.1	1099.7	0.0
0.0	557.4	Cre15.g638000.t1.1	984.6	0.0
1.6 E^{-176}	550.8	Cre15.g638954.t1.1	971.8	0.0
4.5 E^{-180}	549.3	Cre15.g638401.t1.1	913.7	0.0
1.8 E^{-174}	534.6	Cre15.g638304.t1.1	900.2	0.0
6.2 E^{-172}	529.6	Cre15.g638651.t1.1	993.4	0.0
1.7 E^{-172}	528.1	Cre15.g638700.t1.1	954.5	0.0
5.5 E^{-171}	523.9	Cre15.g638050.t1.1	1091.3	0.0
3.4 E^{-168}	520.4	Cre15.g638550.t1.1	902.1	0.0
1.1 E^{-171}	518.9	Cre15.g639650.t1.1	375.9	2.2 E^{-115}
3.8 E^{-165}	513.8	Cre15.g639308.t1.2	906.7	0.0
1.7 E^{-166}	513.5	NCC1	1941.4	0.0
1.1 E^{-163}	511.1	Cre15.g639000.t1.2	1070.5	0.0
5.7 E^{-163}	504.6	Cre15.g639304.t1.2	825.5	0.0
4.4 E^{-159}	492.7	Cre17.g739800.t1.1	847.0	0.0
1.7 E^{-154}	484.6	Cre15.g638650.t1.1	1004.6	0.0
2.8 E^{-151}	478.4	Cre19.g750747.t1.1	900.6	0.0
1.3 E^{-130}	406.4	Cre06.g258051.t1.1	608.2	0.0
5.1 E^{-118}	388.3	Cre15.g638750.t1.2	745.7	0.0
2.8 E^{-65}	238.8	Cre15.g634800.t1.1	224.9	3.3 E^{-60}
5.6 E^{-25}	108.6	Cre15.g639300.t1.1	230.3	1.2 E^{-66}
3.1 E^{-25}	107.8	Cre15.g639614.t1.2	136.3	1.5 E^{-34}
		gene model	Score (/NCC1)	E value (/NCC1)

Gene models (v5.5 annotation) highly similar to either NCC1 or NCC2 are listed in the third column, with the score and E-value obtained in a search using NCC2 as a query shown in the first two columns and those obtained using NCC1 as a query indicated in the last two columns. Genes models located on chromosome 15 are written in black, those located on other chromosomes in Blue. Gene models written in red have a E-value higher than 10^{-100}, but were nevertheless taken into consideration because they are embedded in the OPR cluster on chromosome 15 and have an E-value lower than 10^{-100} with some NCC1/2 paralogs. The more divergent gene model written in green, ova_1_g13706.t1, is located on chromosome 15 but outside of the cluster and corresponds OPR protein most closely related to these paralogs. Called Chlre_OPR68, it was used as an out-group to root the phylogenic tree in Fig. 5D.

Supplemental Material and Methods

Sequence of the synthetic *NCC1^M and NCC2^M* tagged genes for complementation in *Chlamydomonas*

ncc1-HA

GAATTCATGTTACAGTTCGCGCACGGCAGCGCAGTTGGCTTACGCAGAAGCTACAATGCTCCAGCATCAG
CACACATAGGGCTCGAGCTTTGGCCATGCGGCATACGGCTGACCCAGCCTCGCGCCGCTCGGACCGCCGG
GCGCCTCGGCCCCTCGGCGACCGCCGGCCGTGTGGCCCGCGGCCTGGCGTGCAGCTCCGCAACCAGCGGC
CCGGCGGCGGCAGTGCTCAGCGGCGGCGGCCCCCGCCTTGCCGCCACCGCACCTGCTGCCACAGGGGCTG
GCGGCGGTCTGCGGCAGCTACTGCCGCTAGCCGCCTGGGCCCGTCAAGCCGCACCAGCGCCCGCCACTGC
CGCACCTGCAGCAGCCGGGGCGCGCAGCTGTCTCCTGCAGCCGGGGCGGCGCGCGCAGCAGCAGGTGCTG
GCGGCTGCAGGGAGGTACGGCAGTGGCGGCACCAGTCACAGGAGCAGTGGCAGGAGCAGTGACAACAGCA
GCAGCGGACATGGCGGCATCAGCGGCAGCCGCAGTAGCAGTGACGGTGGCGGAAGGGCGGGCAGGGGCAG
GGCAGACGGAGGCCATTGGGCCGGCGGCAGCGGGCGTGGCAGTGGAAGGACGGGCGGTGGTGGTGGCAGG
AGTGGTGGTGGCGGTAGGCCGCGTGACCCCGAAGACGTGATGATCGGGTATGCGACTCTGGAGGAGCTGC
AAGGAGTCATTGATCAGCGGCTGGCGGTGTGGTGCGAGCGGCAAGACGTGAGCACCATGTCAGCTGCCTT
CGGCCGTTGTGGCAAGGTGGGAGGTTGGACTAGCTAGCGGGCAATAAGAAGAGGAGGGGCAGGGGCAGGG
GTCGGGCTTGCCTGCAAGCACGTGCGGGCCTTCCTGCATTGCAGTGCGGGCACTGACAACGCCAACAACA
CACGCACCTCCTTAGCACTTACCAAAGCTCACACGGTACTGCTTAGGGATGCGTTGCATGATGACGGTGG
GGCTGATCCAAGCTTGCTTTTCACCGCCACAGTGAGTGATCGGGGCATCGTCCTTTCCCACTGCCCCCCC
GCCCCCCCGCCCCCCCCCCTGCAGCTAGAGTCAGCACGCGCCGGC~~GGGCC~~GACCGCCACAGCCGCGGCCC
GTGCCAACATCATTGCGGCCCTTGCGCCTGCTCTCCTTCCGCTCGTCCCACGCATCCGGCAGCCGGCCGG
CTGCAGCCTTCCGCTGTTGGCTCTGGCGAAGGCCGGGGCTGCCAGCGACGGGCGTGTGGAGTCACAGCTG
GCGCCGGCACTGCTGCAGCGGCTGGTGGATCCGGTGTTGCTGGATAGTGCCACTCCCCAGGCCCTGGCGA
ACTCACTGTATGCGCTGGGCAAGCTGCGGGAGGATCAGCAGCGGCGGGGCTTGGGCTGGGATCCCACCTC
ATCGCCGCATCTGAAAGCGCTGGCGAGCGCTGTGGCCTCGCGGCTGCGTGCGGCGAGGGGGCACGGCTTC
ACCGCCCAGGGTGTTTCTAACAGCCTGTGGGCGTGCGCCAAGCTGGGGTACCGCGACTCAGCACTCCTGC
TGCCGCTGGCGGAGGCGGCCGCAGCCCTTGCCCCGGACATGAATGCGCAAGACCTGGCCAACAGCCTGTG
GGCGCTGGAGGCTTTGGGCTGCACCAGACCGGAGTTTCGGGCAGTGCTTAAGGCGCTGTGCGGTGCGGCG
CTGCGGCAGCTGCGAACCCCGAAGGAAGCTGCAGCGTTCATTCCGAGACACCTGTCCAACATCTTGCTGG
CGCTGGAGGGGCTGCAGCTGCGAGGCAAGCAGTCCGAGCTGCTGGCGTCGGCTGTGGCTGCAGAGGGTGT
GCGGCGGGGCTTTGCAGGCTTTAACCCGCAAGACCTCAGCAACTCGGCGT~~GGGCAC~~TGGCCAAGATGGGG
TATGGGGCTGGTGCAACGCCGCAAGCCATGGAGCAGAGCCATTGGTATGCGGCCGCGGTGGCAGCAGCTC
AGCGGCCCGGGGTCATGGCCGGCGCCAAGCCGCAGGCCTGGGCCAACCTGCTGTACGCGCTTGCGTTGGT
GCGCCACCAGCCACCGCCCACGCTGCTTGACGCAGGCGCCGCGGCGGCGATGCAGGGAGGCAATGCACAG
GAATGTTCAAACACGCTGTGGGCGCTGGCGTTGCTGCAGCTGCGGCATGCCGGCTTGGAGGCGGCGGTGT
GCGGCCGGCTGGGTGAGCTGCTGCGGCAAGGCCCGGAGTCGCTTATTGCACAGGAGCTTTGCAACAGCCT
GTGGGCGCTGGCAGTGCTGGCAGGCGGCGGTGGCCCTGCTAGCCCGGCTGCCGCTGCGATGGCCCCGGCG
CTTGCCCGCGAAGCCGCGCGGCGCCGTGAGGGGTTTCAGACCGAGGGGCTACTGCAGCTCTGGCAAGCGC
AACATGAGCTGGGCGGCGAGGTGGCGGAGGCTCTTGCCCGCAGCCCCGACCTGCTAGCCGCGATTGAGGC
GGCGGTAGCTGCCGTACAGGCGACGGGGTCGACCACCAGCAGTCTGCAGGAGCAGGTGGCAGAGGCGCTT
CAGCGACTGCTGCAGAAGGGTCGCCTGCCCATCGTTTCGGTGCGGACGGAGGTGGTGGTGGAGGGCGTGC
TGGGGCGTGTGGATATTGTGGCGGACTGGAGCGATGGGCGCAGAGTGGCAATCGAGGTGGACGGGCCAGA
CCACTTCCTCACCAACCGGAAGGACGACCCGTCCGCGGTAAACGGATCCACGGCCCTACGTAACCGGCAG
CTGCAGCGGGCGCTAGGCGAGGGAGGCCTGGTGTGTGTGCCGTTCTGGGAGTGGTACGGTCTGAGGACGC
CTTCCGCTCAGGAGGCCTACCTGCTGCAGCGGCTGCAGGATCTGCTATCGGGCGCTTCGTCAGGAGCAGC
AGGAGGGGGAGGCAGCGCCGCCGCCCCTCGCCACCAGCAGCAGCGGACGACTCGGCTGCAGCCGGACATT
GCCGGCAGCAGCACCACCAGCACCACCAGCACCGCTAACAGCAGCCAGAAGAGGCGCGTCCTCGTTGTCC
GCAGAAAGGCGCTGAAGCAGGAGGCGGCCGGCGGCGGTGGGGCGGGAGGGGCCTCGAATGCAACAGCAGC
TGCAGCGACGGTAGCGGTTGCGGGGGCTGAGTCAGTGGCCCCTGCCTGGCAGCAACCAGTGGAAAAGGCG
CAGCGGCTTCCCAGCAGTGGCGCTGGTGGCGGTGCTGGCGAGGGCGGGGAAGACGGAGCTTCTCTGGCGC
CGCAGCCTCCACCACCGCAGCAGCAG~~CATGTG~~TCAGCGCCCCGGGCGCGCCGCAGCAGCAGCCAAACCCA
AAGCCGTAGCCGCAGTAGCGTGTATACATACCCCTACGACGTTCCGGACTACGCGTACCCGTACGATGTC
CCCGACTACGCTAGCTACCCTTATGATGTTCCTGATTATGCTTGAAGATCT

ncc2-HA

GAATTCATGCGCACCAGCCGGCCGCCTCTCAGCGCGTACAGCAACAGCAGGCATGGCTACGGCTATGGAT
CTGCAGCTGTCAGCCAGGGACCGCGCCCGGCGTCACGCGGGCTGCTGCCTGTTCGCTCGGGCACTACGGT
GCCCAGGGCGTTTGCGAGCCCGAGCTCGGCCTACTCGGCCCGACCTGTCCTGCTGGAGTCGCGCAGCCTC
TCC~~CCACGGAGCGG~~CTCTGTGCCGTCGGTGCTCGCTGGCAGCGGTGGCCGGCGCCCCATCGCCCCCGCCG
TCCCCCCGCCTTCCAGCCGGCAGCTGGAGCAGGCGGCTGCTGGGGGCGGCGGCGCCAGCAGCAGCAGCCG
TGGCGAAGGCCGCATGGGCGGACGCGGCAGAGGGCGTGGTGGCTTCCGGGGCCGTGGAGGCACAGCTGCC
TCGGATGACGAAACGCTGGAGGAGCTGCAAGCGGCCGTGACTTGCCAGCTGCCGGGCTGGGTGGAGCGGG
AGGACACAGCCGCCATCAGCAACGCCTTCCGCAAGGCCGTGCAGGTACGGTGATCGTAGGAAGGCAACGC
CATGATTTGGCTTGTGGCCCTTATGTCTTCATAATAAGCTCACCATGTACGTGCATGCCGCGCTACTCCT
TCCCACATGCAGTATCGCTCGGGCAATGGAGGCGCCGCCACGGCAGCTCGCATCCGCTCCGGCATCATTG
CTGACCTCTCAGCCGCCTATCTGCCCCTGGTGCCGCGCATTCGGAAGCCGTTTGACTGCCGCATGCCGCT
GTGGGCTTTGGGAAAGGCTGGGATTGGCAGCAAGGGCCCAGAGGTGCAGCTGGCGAATGCATTGTTGCAG
CGGCTGCTAGACCCGGCAGTCATTGCGGCGGCAACCGCCATGGACCTGTCTCTTGCGCTTTACGCGCTGG
GCAAGCTGCGTGAAGGCTGGCAGCAGAACGGCGAAGGGTGGGATCAAAGCCTTGGCAAACTGACGGATGC
AATAAAGACCAGGCTCACGGCGGCAGTCGGGCACGGCTTCAACGCGCAGGACGTGTCCAACAGCCTGTGG
GCGTGCGCCAAGCTGGGGCTCGCGGACGCGGACCTGCTGCAGCGACTTGCAGAGGCGGGCGCGGCGGTTG
CTGGGGATATGACACCACAGGCACTCAGCAACAGCCTCTGGGCTCTCGAGGCTCTGGGGTGCACGGGGCC
AGCATACCGGGCAGCGGTTCAGGTACTGTGCGGGGAGGCGCTGCGGCGGCTCCGGACACCCAAGCTCGCT
GAAGCATTCAAGCCACAGGACTTGTCGAACATCCTGCTGGCGCTGGAGGGGCTGCAGCTGGGCAGCGAGC
AGGCACAGTTGGTGTCGGCGGTGGCGGCAGAGGACGTGCGGCGGGGCTTCACAAGGTACAACTCTCAGGA
CATCCGCAACTCCACCTGGGCCCTCGCCAAAATGGGGTTCGGCGTGGGCCCGGAGGCGCCGGCGGAGCAG
CGGCAGTGGGTCACAGCAGCCCTGGACGCAGCTATGCGGCCAGGCACAATGGCGACAGCAACACCACAGA
ATTGGTCCAACCTGCTGTATGCGCTTTCGGTAATGCGCCACCAGCCGCCGCCAGTACTGCTAGACGCGGG
TGCAGCGGCAGCCATGCGCGCCAACAGTGTCAATGGCCCGCAGGACTGCGCGAACACGCTGTACGCGCTG
GCGGTGCTGCAGCTGCGGCATGCCGGTTTGGAGGCGGCGGTGTGCGGCCGGCTGGGCGAGCTGCAGCAGG
AAGATCTGGAGTCACTGACTGAGCAAGGTCTGGCAAACAGCCTATGGGCCGTGGCCGTGTTCGGAGAAAC
CGGTTCTCCGGCCATGCAACAGCTAGCAATGCAGCTGGCACGCGACGCAGCAATCCGTTGGGAGGAGTTT
GCCGACGAGGGCTTGACACAGTTGTGGCAGGCGCAGCAAGCCCTAGGCGGCGAGGTGGCGGCGGCCCTGC
GCGGCAACAGGAGTCTGCAGGCGGCGATGGATAAGGCAGTGGCTACGTACCGGGAAGATACCAAGCACCT
GCCGGATGACCAGAAGCAGCTCCTCGCAGCGCTGCGGCGGCTTGAGCAGCACGGGAGGGAGACAGCAGGA
GGCCTTGCCGTCCAGTCTGTACAAACAGGCGTTGTTGCGCAAGGCGTGTTGACCCCTGTGGATGCGGTGA
TGGGGCTGGTTGACGGGCGGCAGGTGGCAGTGGAGATGCTCGGACCGAAACGCTTCATCTACAACCGTAA
GCAGGACGATCCCACCGCTGTGAACGGCGGTACAGGCATGCGCAATCGGCAGCTGCGGCGGGCGTTTAGC
GAGGGCGGCGTCTTGCTGGTGCCGCACTGGGAGTGGGAGGGCCTGAAGAGCCCGGAGGAGCAGGAGGCCT
ACCTGCTGCTGCGGCTGCAGGAGGTGGCTGCGGCTGCAGAAACCCAGGCGGCGGCGGCAGCAGCTCAGGG
GGCAGGGAAGACCTCAGCTCGCACATCGGCGGCAGGCGCGGCCGCTGCTCCTGCACCTGCACGCCCCGGT
AGCAGCGGTAGCAGTAGCAGCCGCGGCGGCCGTGGCGGCGGCAGTCCTGGCAGCCCCCAGCTGCTGACAC
GGCCCCCGCGGCGGCGCGTGTATACATACCCCTACGACGTTCCGGACTACGCGTACCCGTACGATGTCCC
CGACTACGCTAGCTACCCTTATGATGTTCCTGATTATGCTTGAGGATCC

Sequence of the synthetic *ncc1* and *ncc2* genes.

Restriction sites introduced upstream and downstream of the coding sequence for cloning purposes are written in red; translation initiation and termination codons are boxed; intron sequences are written in purple. The sequence encoding the triple HA is written in blue, with an upstream BstZI restriction site introduced to delete the triple HA tag, if needed, underlined. Nucleotides written in red differ from the wild-type sequence and correspond either to the *ncc1* or *ncc2* mutations (boxed) or to silent mutations introduced to create new restriction sites (underlined), or to delete unwanted restriction sites (crossed out).

References for Supplemental Information:

Bunting, K.A., Roe, S.M., Headley, A., Brown, T., Savva, R., and Pearl, L.H. (2003). Crystal structure of the Escherichia coli dcm very-short-patch DNA repair endonuclease bound to its reaction product-site in a DNA superhelix. Nucleic Acids Res **31,** 1633-1639.

Emanuelsson, O., Nielsen, H., Brunak, S., and von Heijne, G. (2001). Predicting subcellular localization of proteins based on their N- terminal amino acid sequence. J.Mol.Biol.2000.Jul.21;300(4):1005-16. **300,** 1005-1016.

Horton, P., Park, K.J., Obayashi, T., Fujita, N., Harada, H., Adams-Collier, C.J., and Nakai, K. (2007). WoLF PSORT: protein localization predictor. Nucleic Acids Res **35,** W585-587.

Kazrani, A.A., Kowalska, M., Czapinska, H., and Bochtler, M. (2014). Crystal structure of the 5hmC specific endonuclease PvuRts1I. Nucleic Acids Res **42,** 5929-5936.

Kleinknecht, L., Wang, F., Stube, R., Philippar, K., Nickelsen, J., and Bohne, A.V. (2014). RAP, the sole octotricopeptide repeat protein in Arabidopsis, is required for chloroplast 16S rRNA maturation. Plant Cell **26,** 777-787.

Knizewski, L., Kinch, L.N., Grishin, N.V., Rychlewski, L., and Ginalski, K. (2007). Realm of PD-(D/E)XK nuclease superfamily revisited: detection of novel families with modified transitive meta profile searches. BMC structural biology **7,** 40.

Kuras, R., Saint-Marcoux, D., Wollman, F.A., and de Vitry, C. (2007). A specific c-type cytochrome maturation system is required for oxygenic photosynthesis. Proc Natl Acad Sci U S A **104,** 9906-9910.

Lemaire, C., and Wollman, F.A. (1989). The chloroplast ATP synthase in Chlamydomonas reinhardtii. I. Characterization of its nine constitutive subunits. Journal of Biological Chemistry **264,** 10228-10234.

Sievers, F., Wilm, A., Dineen, D., Gibson, T.J., Karplus, K., Li, W., Lopez, R., McWilliam, H., Remmert, M., Soding, J., Thompson, J.D., and Higgins, D.G. (2011). Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. Molecular systems biology **7,** 539.

Small, I., Peeters, N., Legeai, F., and Lurin, C. (2004). Predotar: A tool for rapidly screening proteomes for N-terminal targeting sequences. Proteomics **4,** 1581-1590.

Tardif, M., Atteia, A., Specht, M., Cogne, G., Rolland, N., Brugiere, S., Hippler, M., Ferro, M., Bruley, C., Peltier, G., Vallon, O., and Cournac, L. (2012). PredAlgo: a new subcellular localization prediction tool dedicated to green algae. Mol Biol Evol **29,** 3625-3639.

Taylor, G.K., Heiter, D.F., Pietrokovski, S., and Stoddard, B.L. (2011). Activity, specificity and structure of I-Bth0305I: a representative of a new homing endonuclease family. Nucleic Acids Res **39,** 9705-9719.

Tsutakawa, S.E., Muto, T., Kawate, T., Jingami, H., Kunishima, N., Ariyoshi, M., Kohda, D., Nakagawa, M., and Morikawa, K. (1999). Crystallographic and functional studies of very short patch repair endonuclease. Mol Cell **3,** 621-628.

Zhang, Y. (2008). I-TASSER server for protein 3D structure prediction. BMC Bioinformatics **9,** 40.