Spontaneous dominant mutations in *Chlamydomonas* highlight ongoing evolution by gene diversification.

Boulouis A[1], Drapier D, Razafimanantsoa H, Wostrikoff K, Tourasse NJ[2], Pascal K, Girard-Bascou J, Vallon O, Wollman F-A, Choquet Y*.

UMR 7141, CNRS/UPMC; Institut de Biologie Physico-Chimique, 13, rue Pierre et Marie Curie, F-75005 Paris, France.

present addresses: [1]: Max-Planck Institut of Molecular Plant Physiology, Am Mühlenberg 1, 14476 Potsdam, Germany. [2]: FRC 550, CNRS/UPMC; Institut de Biologie Physico-Chimique, 13, rue Pierre et Marie Curie, F-75005 Paris, France.

*Corresponding author: E-mail: choquet@ibpc.fr

Running title: Dominant mutations in paralogous genes

Keywords: RNA-protein recognition/ mRNA turnover/ helical repeat proteins/ diversifying selection/ *Chlamydomonas reinhardtii*

Character count: 101372

Abstract

We characterised two spontaneous and dominant nuclear mutations in the unicellular alga *Chlamydomonas reinhardtii*, *ncc1* and *ncc2*, which affect two octotricopeptide repeat (OPR) proteins encoded in a cluster of paralogous genes on chromosome 15. Both mutations cause a single amino acid substitution in one OPR repeat and target the mutated NCC1 and NCC2 proteins to the coding sequence of the *atpA* and *petA* chloroplast transcripts respectively, promoting their decay. At variance with the *ncc1* mutation, the *ncc2* mutation requires on-going *petA* mRNA translation to induce its degradation. Using new clues on nucleotide recognition by OPR repeats, we identified the target of the mutated proteins within the *atpA* and *petA* mRNAs. Thus, these mutants undergo a new type of nuclear control on chloroplast mRNAs. The two mutations illustrate how diversifying selection can allow cells to adapt the nuclear control of organelle gene expression to environmental changes. We discuss these data in the wider context of the evolution of the helical repeat proteins-mediated control.

Introduction

In all living beings, the genetic program is mainly carried out, at each step of genome expression, through sequence-specific interactions between nucleic acids and proteins. During and after transcription, RNA-binding proteins control the processing, transport, localisation, translation and stability of coding and non-coding RNAs.

Modular proteins, made up of tandem repeats of simple structural motifs (20–50 AAs in length), most often comprising antiparallel α-helices and, thus, also termed helical repeat proteins, are particularly well-suited to develop interactions with RNAs and other macromolecules. Repeated motifs fold independently and stack on each other to form elongated or concave surfaces, hence assigning the helical repeat proteins to the α-solenoid super family of proteins. While TPR (TetratricoPeptide Repeat: 34 AAs), HEAT (Huntington, Elongation factor 3, protein phosphatase 2A, and yeast kinase TOR1: 39 AAs), Arm (Armadillo repeat: 38 AAs), Ank (Ankirin: 33 AAs) and LRR (Leucine Rich Repeat: 23-24 AAs) motifs are involved in protein-protein interactions, PUF (Pumilio and fem-3 binding factor: 36 AAs), TALE (Transcription Activator Like Effector: 34 AAs), PPR (PentatricoPeptide Repeat: 35 AAs), HAT (Half A Tetratricopeptide repeat: 34 AAs) and mTERF (mitochondrial TERmination Factor: ~30 AAs) motifs mediate protein-nucleic acid interactions (reviewed in (Rubinson and Eichman, 2012)). Recent crystallographic structures of PUF (Wang et al., 2001; Wang et al., 2002; Miller et al., 2008), TALE (Deng et al., 2012; Mak et al., 2012) and PPR (Ke et al., 2013; Yin et al., 2013) proteins in complex with their RNA/DNA target confirmed that nucleic acids bind in an extended conformation to the inner concave surface of the solenoid, with each nucleotide contacting one -at most two- consecutive repeats. Repeats, thus, act in a modular fashion, each repeat interacting with one nucleotide. Within a repeat, the side-chain of a few amino acids at specific positions determines the recognised nucleotide, mainly by establishing hydrogen bonds with the nucleotide base. Prediction of the nucleotide recognised by specific amino acid combinations has recently been made successfully for PUF (Wang et al., 2002; Cheong and Hall, 2006; Filipovska et al., 2011), TALE (Boch et al., 2009; Moscou and Bogdanove, 2009) and PPR (Barkan et al., 2012; Yagi et al., 2013) proteins. Based on this "recognition code", recombinant PUF and TALE proteins can be engineered to bind virtually any RNA or DNA target of interest (Christian et al., 2010; Cooke et al., 2011), for reviews: (Bogdanove and Voytas, 2011; Filipovska and Rackham, 2011, 2012; Yagi et al., 2014). Furthermore, this modular architecture endows helical repeat proteins with a great versatility *in vivo*, as module

reorganisation through genetic recombination or substitutions of nucleotide-specifying amino-acids will confer specificity for new target sequences.

Nuclear control of organelle gene expression is a key feature of eukaryotic cells that emerged after endosymbiosis (Choquet and Wollman, 2002; Woodson and Chory, 2008; Barkan, 2011). Indeed, every post-transcriptional step of organelle gene expression, either editing, splicing, processing from polycistronic to monocistronic transcripts, 5' or 3' end trimming or translation activation, is controlled, in a gene- hence sequence-specific manner by nucleus-encoded RNA-binding proteins (denoted ROGEs for Regulators of Organelle Gene Expression, for reviews see (Barkan and Goldschmidt-Clermont, 2000; Choquet and Wollman, 2002; Schmitz-Linneweber and Small, 2008; Woodson and Chory, 2008; Germain et al., 2013; Barkan and Small, 2014). Predictably, most ROGEs belong to helical repeat protein families (PPR, HAT, mTERF, reviewed in: (Barkan and Small, 2014; Hammani et al., 2014)). The great expansion of ROGEs in all photosynthetic organisms contrasts with their poor conservation between different lineages. Indeed, despite their common structural organisation, the various families of helical repeat proteins do not generally share a common origin as their respective consensus motifs are not related. This suggests a high flexibility of nucleo-organelle interactions, well-suited for a rapid adaptation to new environmental constraints or ecological niches. For instance, PPR proteins, predominantly targeted to mitochondria or chloroplasts (Lurin et al., 2004) are particularly numerous in land plants with more than 500 members identified in *Arabidopsis* or rice. By contrast, the unicellular green alga *Chlamydomonas reinhardtii*, hereafter referred to as *Chlamydomonas*, possesses only 14 PPR proteins (Tourasse et al., 2013) but more than 120 members of another family of helical repeat-proteins, poorly represented in land plants, the OPR (OctatricoPeptide Repeats) proteins defined by a degenerated motif of 38 residues (Loiselay, 2007; Eberhard et al., 2011; Rahire et al., 2012). As PPR repeats, OPR repeats are predicted to fold into a pair of antiparallel α-helices. Most OPR proteins are predicted to be targeted to organelles (Loiselay, 2007) where several have been shown to control the post-transcriptional steps of gene expression.

Most mutants affected in ROGEs described to date were screened for photosynthetic or respiratory defects after mutagenesis. They display a recessive phenotype, being defective in a gene product that, in the wild-type, binds specifically to a given target transcript, usually in its 5'UTR. In contrast, the two nuclear mutations that we describe below in *Chlamydomonas* appeared spontaneously and are dominant. They correspond to single amino-acid substitutions in two OPR proteins that gain a new function by recognising a new target in

the coding region of two chloroplast transcripts, thus providing insights into the evolution of the nuclear control of organelle gene expression.

Results

Isolation of the *ncc2* mutation, which alters the stability of the *petA* transcript.

The 5'*petA-psbB* chimera is made of the *psbB* coding sequence, encoding the PSII core antenna CP47, translated under the control of the 5'UTR of the chloroplast *petA* gene, encoding the cytochrome *f* subunit of the cytochrome b_6f complex (Fig. 1A). When inserted by chloroplast transformation at the *petB* locus, this chimera only allows a low level of CP47 expression, insufficient to sustain the phototrophic growth of {5'*cA*CP47} transformed cells, which display a PSII defective phenotype (Fig. 1 B-C). While plating the {5'*cA*CP47} strain on Minimum medium, one spontaneous phenotypic revertant, Su0-T9, appeared which had recovered phototrophic growth capability and increased accumulation of CP47 (Fig. 1 D-E). In crosses to a wild-type strain (*mt*-), all tetrad progeny inherited the 5'*petA-psbB* chimera, uniparentally transmitted by the *mt*+ parent, but the photoautotrophic capability segregated 2:2. This was indicative of a single nuclear suppressor mutation that we called *ncc2* for **n**uclear **c**ontrol on **c**hloroplast gene expression. We then crossed one phototrophic progeny (*mt*-) to the wild type (*mt*+) to eliminate the 5'*petA-psbB* chimera as all tetrad members inherited the wild-type chloroplast genome of the *mt*+ parental strain (see Table I for strains generated by crosses in the course of this work): two tetrads progeny had a wild type phenotype, while the other two progeny, presumably carrying the *ncc2* mutation, were able of photoautotrophic growth but presented fluorescence induction kinetics typical of a cytochrome b_6f leaky mutant (Fig. 2A). Indeed, their Φ_{PSII} (0.2), much lower than that of the wild-type (0.59), indicates a decreased electron flow downstream of PSII (Maxwell and Johnson, 2000). This correlates well with the 6-fold reduced abundance in cytochrome (cytochrome) *f*, when compared to the wild-type strain (Fig. 2B). In 5 min pulse labelling experiments, the rate of synthesis of cytochrome *f* was much reduced and hardly detectable in the *ncc2* mutant (Fig. 2C). When probing chloroplast transcripts for the major photosynthetic protein subunits by mRNA-hybridisation, we observed a selective drop in accumulation of the mature *petA* mRNA (below 5% of the wild-type level: Fig. 2D) whereas *petA* processing intermediates were little affected (stars on Fig. 2D, top panel). The latter observation suggests that, in the mutant, transcription of the *petA* gene is preserved, while the mature *petA* transcript undergoes accelerated degradation.

The latter phenotype readily explains the suppressor effect of the *ncc2* mutation on the expression of the 5'*petA-psbB* chimera. Indeed, cytochrome *f* is a CES protein, whose rate of synthesis is controlled by its rate of assembly within the cytochrome b_6f ((Choquet et al.,

1998; Choquet et al., 2003), reviewed in (Choquet and Wollman, 2008)). In the wild type, a small fraction of cytochrome *f* remains unassembled and down-regulates three-fold the rate of translation of the *petA* mRNA -or of any chimera driven by the *petA* 5'UTR, such as the 5'*petA-psbB* chimera here-. The *ncc2* mutant synthesizes reduced amounts of cytochrome *f* (Fig. 2C), which precludes the accumulation of significant amounts of unassembled cytochrome *f*, thereby releasing down-regulation, which leads to increased rates of translation of 5'*petA*-driven genes (Choquet et al., 1998; Choquet et al., 2003). The 5'*petA-psbB* chimera, now expressed at a much higher rate in the *ncc2* mutant than in a wild-type nuclear background, can drive synthesis of CP47 at rates high enough to sustain photoautotrophic growth, even though still lower than those of the endogenous *psbB* gene.

We previously identified two nuclear genes, *MCA1* and *TCA1*, which contribute to the protection of the *petA* transcript against exonucleolytic degradation (Wostrikoff et al., 2001; Raynaud et al., 2007; Loiselay et al., 2008; Boulouis et al., 2011). However, crosses of *ncc2* to either *mca1-2* or *tca1-2* mutants (Supplemental Table SI) clearly showed absence of linkage, indicating that the *ncc2* phenotype was not due to altered expression of *MCA1* or *TCA1*.

The *ncc2* mutation is a dominant mutation.

To determine whether the *ncc2* mutation was recessive or dominant, we constructed vegetative diploid strains either homozygous or heterozygous for the *ncc2* locus, as described in Table I. While homozygous *NCC2/NCC2* diploids were indistinguishable from wild-type, heterozygous *ncc2*/*NCC2* diploids, as well as *ncc2/ncc2* homozygotes displayed the same 20-fold reduced accumulation of the *petA* mRNA as the haploid *ncc2* parental strain (Fig. 2E). Some gene-dosage effect, however, partially damped the effect of the *ncc2* mutation at the protein level, since cytochrome *f* abundance only decreased two-fold in the *ncc2/NCC2* heterozygote (Fig. 2E). Thus, in contrast to most mutations in ROGE genes described to date in *Chlamydomonas*, the *ncc2* mutation is not recessive but dominant. In that respect it resembles the dominant *ncc1* mutation that reduces the accumulation of monocistronic transcripts of the chloroplast *atpA* gene (Drapier et al., 2002). The dominant *ncc2* mutation is thus unlikely to be a loss-of-function allele, but would rather modify some gene product, now acting on the *petA* mRNA.

The *ncc2* mutation acts on the *petA* coding sequence, as does the *ncc1* mutation on the *atpA* coding sequence.

In *Chlamydomonas* chloroplasts, the ROGEs controlling the stability of a chloroplast transcript identified to date target its 5' UTR, with the exception of the factor identified by the *ncc1* mutation, which targets the *atpA* coding sequence (Drapier et al., 2002). We thus tested which part of the *petA* mRNA was targeted by the *ncc2*-dependent degradation process, using two distinct *petA* chimera.

At the *petA* locus, we first substituted the sequence coding for mature cytochrome *f* by that encoding the ATP synthase subunit α, truncated to maintain a similar mRNA length, fused in frame with the sequence coding for the lumen targeting peptide of cytochrome *f* ($\Delta f::\alpha_{Tr}$ chimera, Fig. 3A, see Table II for the list of transformants generated during this work). As for all chloroplast transformations in this study, the recipient strain was of mt^+ mating type and the transforming plasmid also carried an *aadA* cassette conferring resistance to spectinomycin to allow the selection of transformants. One of the resulting transformants, $\{\Delta f::\alpha_{Tr}\}$, was crossed to the *ncc2* mt^- mutant. All tetrad offspring inherited the chloroplast chimeric transgene, but only half progeny inherited the *ncc2* nuclear mutation. RNA-blot analysis showed no hybridisation with an intragenic *petA* probe in the parental strain $\{\Delta f::\alpha_{Tr}\}$, because this sequence had been replaced by that of *atpA*. An *atpA* probe detected, in addition to the endogenous *atpA* transcripts, the $\Delta f::\alpha_{Tr}$ chimeric transcript in the parental transformant and in all progeny (Fig. 3B). This was also true for a larger and minor transcript (indicated by *), resulting from co-transcription of the chimera with the downstream *aadA* cassette. Strikingly, these two bands accumulated to the same extent in all progeny, irrespective of their *ncc2* or *NCC2* genotype. Thus, the *ncc2* mutation did not decrease the accumulation of the chimeric transcript: the sequence coding for mature cytochrome *f* is required for *ncc2*-mediated decay of the *petA* transcript.

Symmetrically, the *petA* coding region, fused to the *atpA* 5'UTR and *rbcL* 3'UTR in the *dAfR* chimera (Fig.3A), was introduced by chloroplast transformation in a wild-type strain. After crossing with *ncc2 mt-*, all tetrad progeny expressed the chimeric *petA* transcript, larger than the genuine *petA* transcript. However, two members of each tetrad, likely carrying the *ncc2* allele, showed a markedly reduced accumulation of the chimeric transcript as did the endogenous *petA* mRNA in the *ncc2* parent (Fig. 3C). The *petA* coding sequence is thus not only required but also sufficient to make a transcript sensitive to the *ncc2* mutation.

Active translation of the target transcript is required for its destabilisation in *ncc2*, but not *ncc1*, backgrounds.

Since the *ncc1* and *ncc2* mutations target respectively the *atpA* and *petA* coding sequences, we wondered whether translation of these transcripts was required for their degradation in the *ncc1* and *ncc2* mutated contexts. We used the *mt*+ {$atpA_{Stop}$} and *mt*+ {$petA_{Stop}$} strains, which express untranslatable *atpA* and *petA* mRNAs because their initiation codons are replaced by a stop codon (Boulouis et al., 2011; Eberhard et al., 2011) and crossed them with the *ncc1* and *ncc2* mutants (*mt*-), respectively. The transcript pattern for *atpA* in the {$atpA_{Stop}$} mutant is similar to that of the wild-type *atpA* gene, which comprises four mRNAs transcribed from the tetracistronic *atpA* gene cluster (Drapier et al., 1998). The *ncc1* mutation markedly decreases the amount of monocistronic *atpA* transcript relative to the polycistronic forms (Drapier et al., 2002). A 2/2 segregation in the ratio of di- to mono-cistronic *atpA* transcripts abundance was clearly visible in the progeny from the {$atpA_{Stop}$} x *ncc1* cross, with progeny 1 and 2 being *ncc1*-like in the tetrad shown in Fig. 4A. Thus, the destabilising effect of the *ncc1* mutation is still observed when *atpA* transcripts are not translated. In contrast, the *ncc2* mutation had no effect on the $petA_{Stop}$ transcript, whose accumulation remained high and similar to that in the {$petA_{Stop}$} parent for all tetrad progeny (Fig. 4B). Destabilisation of the *petA* transcript by the *ncc2* mutation likely depends on active translation.

Still, this result could indicate that the AUG initiation codon is part of the recognition site of the mutated NCC2 factor on the *petA* transcript. To gather independent evidence for the contrasting behaviours of the *ncc1* and *ncc2* mutations towards translation of their target transcripts, we studied transcript accumulation over time in cultures treated with lincomycin, an inhibitor of chloroplast translation. The drug did not affect the accumulation of *atpA* transcripts in strain *ncc1* (Fig. 4C). By contrast, the accumulation of the *petA* transcripts in the *ncc2* mutant increased spectacularly over four hours of incubation with lincomycin, from barely detectable before addition of the translation inhibitor, up to wild-type levels after 4 hours (Fig. 4D). Thus, active translation of *petA* transcripts is required for their *ncc2*-dependent destabilisation, whereas the *ncc1* mutation destabilises the *atpA* transcript independent of translation.

Towards a more accurate localisation of the targets of the *ncc2* and *ncc1* mutations

To better understand how *petA* mRNA translation triggers its degradation in the *ncc2* mutated context, frame-shifts were introduced within the *petA* coding sequence after nts 93 or 390, causing premature translation abortion after codons 42 and 145 respectively (Fig. 5A). The f_{42}St mutation preserves the translation of the lumen targeting peptide, which is not

sufficient to confer *ncc2*-sensitivity to a translated sequence (Fig. 3C). After chloroplast transformation in a wild-type strain, neither of these constructs enabled detectable accumulation of a truncated cytochrome *f* in the resulting transformants {f_{42}St} and {f_{145}St}. Tetrad analysis after crossing these transformants to a *ncc2 mt-* strain, showed that the *ncc2* mutation still decreased the abundance of the f_{145}St transcript 20-fold (Fig. 5B). In contrast, transcripts from the f_{42}St construct were much less sensitive to the *ncc2* mutation as their abundance only decreased two-fold in *ncc2* progeny (Fig. 5C). Thus, the *ncc2*-mediated degradation of *petA* transcripts starts early after the lumen targeting peptide has been synthesised and is completed before translation reaches the 145th codon of cytochrome *f*.

The target of the NCC1 mutated factor was previously localised in the last 1360 bases of the *atpA* mRNA (Drapier et al, 2002). We narrowed down this region by transforming the chloroplast of the *ncc1* strain with the Δ*f*::α_{Tr} construct described in Fig. 3A, which contains the *atpA* coding sequence deprived of its last 579 nucleotides. In contrast to the endogenous *atpA* transcript, the chimeric Δ*f*::α_{Tr} mRNA was not destabilised in the *ncc1* background (Fig. 5D), which points to a localisation of the NCC1 target within the last 579 nucleotides of the *atpA* transcript.

Identification of the *ncc2* and *ncc1* mutations

The *ncc1* and *ncc2* mutations being dominant and still allowing photoautotrophic growth, we could not use complementation to clone the genes. Thus, we mapped the mutations onto the nuclear genome of *C. reinhardtii* by crossing the two mutants with the S1D2 strain (*mt*$^-$), which shows a profusion of polymorphism with *C. reinhardtii* laboratory strains (Gross et al., 1988) and has been extensively used for molecular mapping (Kathir et al., 2003; Rymarquis et al., 2005). We harvested about 50 tetrads from each cross and picked one mutant per tetrad. The *ncc2* progeny were identified based on their fluorescence phenotype. To identify *ncc1* progeny, we used for the cross the non-photosynthetic strain *ncc1* *mt*$^+$ {FAFA}, bearing the coding sequence of *atpA* fused to the 5' and 3' UTRs of the *petA* gene. This chimera is expressed at a low level in a wild-type background and, when combined with the *ncc1* mutation, prevents phototrophic growth (Drapier et al., 2002). By PCR-based mapping with diagnostic markers on all chromosome arms (Kathir et al., 2003), we found linkage for both mutations to the *ZYS3* marker on the long arm of chromosome 15 (96 % and 71% for the *ncc2* and *ncc1* mutations, respectively, Fig. 6A). Using new markers listed in Suppl. Table SII online, the position of the *ncc2* mutation was restricted to a DNA region

between nucleotides 659,176 and 1,063,367 (*Chlamydomonas* genome version 5.5, available on Phytozome at: http://phytozome.jgi.doe.gov/ (Merchant et al., 2007; Blaby et al., 2014))

We then crossed strains *ncc2 mt-* and *ncc1 mt+* {FAFA} and recovered a few double mutants among 80 tetrads - identified both by their *ncc2*-like fluorescence induction kinetics and by their poor phototrophic growth, due to the *ncc1* {FAFA} combination. Double mutants showed decreased levels of both *atpA* and *petA* transcripts, which were respectively comparable to those in the parental *ncc1* {FAFA} and *ncc2* strains. Accordingly, the accumulation of both subunit α and cytochrome *f* was decreased, as shown on Supplemental Fig. S1 online for one typical double mutant. Its genome was sequenced using the Illumina platform (2 x 100 bp). After eliminating (see Supplementary methods) synonymous and intergenic polymorphisms and those also present in a collection of quasi-isogenic photosynthetic mutants carrying neither *ncc1* nor *ncc2* mutations, we were left with two point mutations in the interval determined by molecular mapping: an A → C substitution at position 693,478 and a double substitution AT → GG at position 1,001,095/6. Each of these mutations causes a single amino acid substitution in two distinct genes encoding OPR proteins: Cre15.g638950.t1 and Cre15.g640400.t1. The structure of these two gene models is fully supported by EST evidence (Sanger, 454 and Illumina data displayed at http://genomes.mcdb.ucla.edu/Cre454/).

Direct sequencing of specific PCR products amplified from wild-type, *ncc1, ncc2* and *ncc1 ncc2*{FAFA} strains (for primers see Table EIII online) confirmed the absence of these mutations in the wild-type and the presence of both in the double mutant. The A → C substitution was found in the genome of the *ncc2* mutant, but not in that of the *ncc1* mutant, and leads to the S_{431} → R substitution in Cre15.g640400.t1, which was therefore named *NCC2*. Conversely, the *ncc1* mutant, but not the *ncc2* mutant, carried the AT → GG mutation, leading to the D_{568} → A substitution in Cre15.g638950.t1, thereafter called *NCC1*.

The NCC1 and NCC2 proteins both contain 9 OPR repeats, the mutations changing the 6^{th} residue of the 6^{th} OPR repeat of NCC1 and the 8^{th} residue of the 6^{th} OPR repeat in the NCC2 protein (Fig. 6B-C). In addition, both proteins also contain a ~60 residues RAP (RNA-Binding Abundant in Apicomplexa; (Lee and Hong, 2004)) domain at their C-terminus. The central regions of NCC1 and NCC2, containing the OPR repeats and the RAP domain, are highly similar (57 identity, 69% similarity), while their N and C-terminal regions are more divergent.

Transgenic expression of the mutated NCC1(/NCC2) protein confers the *ncc1*(/*ncc2*) phenotype to transformed strains

The *ncc1* and *ncc2* mutations being dominant, we predicted that mutated copies of the *NCC1/NCC2* genes, when introduced by transformation into the wild-type strain, should confer the *ncc1/ncc2* phenotypes to the transformed cells. The sequences coding for the mutated NCC1 and NCC2 proteins (hereafter referred to as $NCC1^M$ and $NCC2^M$), including the intron, were fused to a triple HA tag to allow their immunodetection in transformed cells, and introduced into *Chlamydomonas* by transformation, using a vector carrying also the *aphVIII* paromomycin resistance gene (Sizova et al., 1996; Sizova et al., 2001). Clones, selected on paromomycin-supplemented TAP plates, were screened with an HA-specific antibody for expression of $NCC1^M$ or $NCC2^M$. As predicted from their respective molecular mass, $NCC2^M$ migrates slightly faster than $NCC1^M$ (Fig. 7A). We then assessed the accumulation of the *atpA/petA* transcripts by RNA blots. Transformants expressing $NCC1^M$ showed a marked decrease in the accumulation of the *atpA* monocistronic transcript, resulting in an inversion of the mono- to di-cistronic ratio, typical of *ncc1* mutants. When independent transformants were compared, we found that the higher the accumulation of $NCC1^M$, the less *atpA* monocistronic transcripts (Fig. 7B), consistent with an $NCC1^M$–mediated degradation of the *atpA* mRNA. Similarly, the accumulation of the *petA* mRNA was strongly reduced in transformants expressing $NCC2^M$ (Fig. 7A), even if we never obtained levels as low as in the *ncc2* strain, probably because of insufficient expression level of the transgenic $NCC2^M$. When comparing two transformants expressing $NCC2^M$, the extent of *petA* mRNA destabilisation correlated with the abundance of the protein (Fig. 7B).

Identification of the $NCC1^M$ and $NCC2^M$ targets.

Based on a preliminary version of the code for nucleotide recognition by OPR repeats (Manuscript in preparation), we identified the $NCC1^M$ and $NCC2^M$ target sequences as NAGNGATTA and GTGAGGNTA. These sequences were indeed found at positions 1066-1075 and 130-139 of the *atpA* and *petA* coding sequences, respectively. It is of note that the latter sequence is located five bp downstream of the premature stop codon in the chimera f_{42}St, a region that we identified above as critical for the degradation of the *petA* transcript in the *ncc2* mutant (Fig. 5C).

To confirm the assignment of the target sequences, we first assessed the predicted target of $NCC1^M$ by introducing a unique change at the position of the nucleotide presumably recognised by the mutated OPR repeat. We chose a nucleotide substitution that leads to a

conservative substitution $I_{391} \rightarrow V$ at the protein level (Fig. 8A). After transformation of the wild-type and *ncc1* strains, the mutated *atpA* gene, *atpAT1*^{M}, replaced the endogenous *atpA* gene. In the wild-type background, this mutation caused a limited decrease (~20%) in the accumulation of the monocistronic *atpA* transcript and a slight increase in the accumulation of the dicistronic *atpA* transcript (Fig. 8B). When expressed in the *ncc1* background, the mutation in contrast increased the level of *atpA* transcripts, with a fourfold increase in monocistronic transcript and a twofold increase in dicistronic transcript. As a result the Mono/Di ratio was higher than in the original *ncc1* strain, as expected, although not as high as in the wild-type. Thus, the ability of NCC1^{M} to destabilize the *atpA* transcripts was affected, but not abolished, by a point mutation in the predicted target.

The partial suppression of the NCC1^{M} effect, when only one nucleotide substitution was introduced in its target, prompted us to introduce larger changes when assessing the target sequence of NCC2^{M}. We introduced several silent mutations in the predicted *petA* target, converting the original GTGAGGCTA sequence to G**A**GA**A**GC**A**A (Fig. 9A). After biolistic transformation of the wild-type and *ncc2* strains we observed that the mutation *petAT2*^{M}, had no effect on the *petA* mRNA steady state level in a wild-type genetic context. In stark contrast, it led to a spectacular restoration of the accumulation of *petA* mRNA in the *ncc2* strain, up to wild-type level (Fig. 9B), thus completely abolishing the *ncc2* phenotype, demonstrating that we indeed identified the target of NCC2^{M}.

We then wondered whether the sequence of the NCC2^{M} target would still be recognised in a different nucleotide context, that is when placed in an unrelated chloroplast mRNA. We thus inserted the target of NCC2^{M}, either in the *petD* 5'UTR (5'*petD*::*T2*,) or within the *petD* coding sequence (*petD*$_{CDS}$::*T2*), taking care to preserve the *petD* translation frame (Fig. 9C). We used these two insertion sites to assess whether active translation would still be required for the destabilisation of the *petD* transcript. When transformed with the 5'*petD*::T2 chimera both wild-type and *ncc2* strains remained phototrophic and the mutated *petD* transcript accumulated to the same level as does the original *petD* transcript in the wild-type (Fig. 9D), as previously found for other mutations at this neutral site (Sakamoto et al., 1994; Higgs et al., 1999). In contrast, chloroplast transformants expressing the *petD*$_{CDS}$::T2 chimera were incapable of phototrophic growth, neither in wild-type nor in *ncc2* nuclear backgrounds, as expected from the probable impairment of transmembrane helix II insertion within the membrane. In the *ncc2* background, the accumulation of the chimeric *petD* mRNA was largely prevented whereas it remained unaffected in the wild-type background, when compared to that of a wild-type *petD* gene (Fig. 9D). Moreover, a 4 hr lincomycin treatment

of the *ncc2* {$petD_{CDS}$::T2} strain restored the accumulation of the chimeric *petD* mRNA to a wild-type level, confirming the dependence of the *ncc2* phenotype on the translation of its target (Fig. 9E). Because the amino acids synthesised from the $NCC2^{M}$ target in the *petD* chimera differ from those encoded by the *petA* gene (Fig. 9C), translation *per se*, rather than the sequence of the nascent peptide, is required for the action of $NCC2^{M}$ on its target transcript.

NCC1 and _NCC2_ belong to the _NCL_ subfamily of paralogs encoding OPR-RAP proteins.

In BLAST searches against the *Chlamydomonas* proteome, the NCC1 and NCC2 proteins hit each other with low E-values (< 10^{-100}), along with a set of 36 closely related proteins (table EIV). The similarity between these proteins was much higher than with any other protein (E>10^{-30}) in the non-redundant Protein database at NCBI (http://www.ncbi.nlm.nih.gov/protein/), including those of the closely related alga *Volvox carteri*. Thus, *NCC1* and *NCC2* are part of a *Chlamydomonas*-specific group of 38 highly similar paralogous genes that will hereafter be called *NCL*, for *NCC-Like* (described in Supplemental Table EV). As shown on the alignment (Suppl. Fig. S2A online), NCL proteins comprise a highly conserved central region, containing 7-12 OPR repeats. 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). OPR repeats are followed by a C-terminal RAP domain (RNA-binding domain Abundant in Apicomplexans, a ~60-residue domain found in proteins of almost all eukaryotic lineages, proposed to be a RNA-binding domain (Lee and Hong, 2004)). By contrast, N- and C-terminal extensions, upstream and downstream of this conserved block, are more divergent.

Strikingly, 32 *NCL* genes, out of 38, are clustered on the long arm of chromosome 15 between positions 686,690 and 1,113,927 (Fig. 6A), while 3 are found on scaffold 19, not yet assembled within a chromosome and 3 isolated genes lie on chromosomes 4, 6 and 17.

At variance with the bulk of OPR-encoding genes (mean size of 1302 codons), which have an average number of 12 introns regularly scattered along the coding sequence, most *NCL* genes (mean size of 806 codons) possess a single intron within the coding sequence, at a conserved position with respect to the protein sequence (Table SV, last two columns and Fig. S2A online). Most probably this intron was already present in an ancestor gene that gave rise to the whole *NCL* subfamily by duplications. Phylogeny of NCL proteins was studied using

Maximum Likelihood inference (Fig. 6D) and was well correlated with the position and orientation of *NCL* genes along the cluster (Fig. 6A, D), suggesting that local tandem gene duplications played a major role in the expansion of the *NCL* family. This evolution is probably still going on: for example, the adjacent *NCL7*, *NCL8* and *NCL9* genes probably originated from very recent duplication events, as *NCL7* and *NCL8* only differ from *NCL9* by 6 and 15 bases, respectively, along the 2277 bp coding sequence. This leads to three amino-acids substitutions in NCL8 and to very limited changes in the very C-terminal end of NCL7. Conversely, the *NCL* subfamily contains at least 2 inactivated genes, *NCL2* and *NCL21*, whose sequences, although similar to that of other *NCL* genes over 598 and 1171 codons respectively, are interrupted by premature Stop codons at positions 200 and 305. According to the expression data available on Phytozome, *NCL2* does not seem to be expressed at the RNA level. The *NCL* subfamily also comprises truncated genes, such as *NCL12* and *NCL16*, that only encode the N-terminal extension and the first four OPR repeats, hence explaining their higher BLAST E-value in Table SIII. Thus, the NCL genes appear to be actively evolving through a “birth and death” process.

NCL genes evolve under diversifying selection pressure.

Within the conserved block, successive OPR repeats, although all obeying the OPR consensus, significantly differ one from another (Fig. S2B). By contrast, the sequence of a given repeat is remarkably conserved between the different NCL proteins (Fig. S2A), except at some variable positions (e.g. 3 and 6), suggesting that the NCL proteins may be under diversifying selection. Diversification was also suggested by the spontaneous appearance of the *ncc1* and *ncc2* mutations. We further tested this hypothesis by computing the nonsynonymous vs. synonymous substitution ratio (ω=dN/dS) for NCL proteins. After mutations, non-synonymous substitutions, when detrimental to the function of the protein are counter-selected, a process called purifying (or "negative") selection (ω<1). At some sites, however, non-synonymous substitutions can be selected for (ω>1), if new protein sequences provide enhanced fitness to the organism (diversifying or "positive" selection). To test whether *NCL* genes were submitted to diversifying selection, we used the program suite PAML which reconstitutes the evolution of codons based on an alignment. dN/dS ratios were compared with predictions of evolution models allowing the presence of sites with ω>1 (diversifying selection), or not (purifying or neutral selection) (Yang and Swanson, 2002; Yang, 2007). Likelihood ratio tests shown in Table III clearly show that models allowing a

class with ω>1 much better fit the observed values, as judged based on χ^2 test p-values. Thus, *NCL* genes are under diversifying selection, which likely generates new RNA-binding specificity via changes in OPR repeats. In the alignment, 21 sites were considered to be under diversifying selection at 99% confidence (37 at 95%), most of them within OPR repeats.

In conclusion, *NCL* genes are still under dynamic evolution and undergo a "birth, diversification and death" process that is driven by diversifying selection pressure.

Discussion

ncc1 and *ncc2*, two dominant and atypic mutations in two OPR proteins, leading to RNA degradation rather than RNA protection.

All mutations affecting ROGEs characterised so far in *Chlamydomonas* are recessive, as they inactivate a protein that binds specifically to a given target transcript, usually in its 5'UTR. Thus, ROGEs likely coevolved with the 5'UTR of their target mRNA. By contrast, the *ncc1* and *ncc2* mutations described here represent a completely new category of ROGE mutations. Both appeared spontaneously in our laboratory strains, both are dominant and both target the coding sequence of their target mRNA. Their unusual dominant nature results from their molecular basis: a single amino acid substitution in an OPR repeat of two different OPR proteins. By introducing a mutant copy of either *NCC1* or *NCC2* into the wild-type strain, we show that indeed these substitutions are sufficient to destabilize the *atpA* or *petA* transcripts. These substitutions in an OPR repeat would change the recognised nucleotide and create new targets for the mutated proteins that fortuitously lie within a coding sequence. As observed for PPR proteins and in agreement with our preliminary version of the OPR code (Manuscript in preparation), well-conserved positions (e.g. 4, 5, 7-15, see fig E3) within the first antiparallel α-helix of the OPR repeats should mostly contribute to protein scaffolding whereas more variable positions (e.g. 3 and 6) would be involved in nucleotide recognition. The *ncc1* mutation, that changes the variable 6th residue of the repeat, was thus expected to alter nucleotide recognition. Surprisingly, although the *ncc2* mutation changes the quite conserved 8th residue, it nevertheless leads to the recognition of a new target.

ROGEs that bind the 5'UTR of their target transcript either activate its translation or protect it from exonucleolytic degradation. In contrast, the interaction of NCC1^{M} and NCC2^{M} with the *atpA* and *petA* coding sequences leads to their degradation by a mechanism that, however, remains to be studied: NCC1^{M} and NCC2^{M} may recruit an endonuclease, or may themselves carry an endonucleolytic activity. In the latter case, the RAP domain found at the C-terminus of both proteins could be responsible for this function, as structural modelling of the RAP domain of NCC1 and NCC2 by the I-TASSER software (Zhang, 2008) used four endonucleases as major templates (see Suppl. Figs. E4 and E5 online). Interestingly, the sole OPR protein in *Arabidopsis*, required for the proper processing of *rrnS* precursor (Kleinknecht et al., 2014), also contains a RAP domain. The other NCL proteins, assuming

that they have acquired some targets within organelle transcripts, could thus also be involved in some processing events.

A link with translation.

A striking feature of NCC2^{M} is that it degrades its target transcript only upon translation. This was observed in two different sequence contexts, within the *petA* and *petD* transcripts, in which the target is likely involved in widely different secondary structures. It is therefore unlikely that NCC2^{M} binding to its target site requires the ribosome-mediated unfolding of RNA secondary structures. Rather, the degradation of the transcript would depend on a tight contact between the ribosome and NCC2^{M}, as supported by the limited decrease in *petA* mRNA that we observed upon early translation termination, a few nucleotides upstream of the NCC2^{M} target. This interaction could change the conformation of either NCC2^{M} itself or the interacting endonuclease, thereby activating nucleolytic activity, as shown for the ribosome associated YeoBI, RelE and RegB RNases (Odaert et al., 2007; Neubauer et al., 2009; Feng et al., 2013).

By contrast, the degradation of the *atpA* transcript by NCC1^{M} does not require ongoing *atpA* translation. However, the monocistronic *atpA* transcript is by far more targeted to degradation than polycistronic *atpA* transcripts. Possibly, the target of NCC1^{M}, localised 460 bp before the end of the *atpA* coding sequence, is only fully accessible in the monocistronic *atpA* transcript, being trapped within secondary structures with downstream sequences in polycistronic transcripts.

NCC1 and NCC2 belong to the NCL subfamily of paralogous OPR-RAP proteins.

NCC1 and NCC2 belong to the NCL subfamily of highly similar OPR proteins. *NCL* genes differ in many respects from the bulk of genes encoding OPR proteins in *Chlamydomonas* (hereafter "*OPR* genes"). Unlike most *OPR* genes, for which orthologs can be easily found in *V. carteri*, probably because of a conserved role in organelle biogenesis, *NCL* genes are specific to *Chlamydomonas*, and thus probably appeared after its separation from the *Volvox* lineage. They share a single intron at a fixed position in the gene and show high sequence similarity, indicating recent appearance by gene duplication. Their genomic organisation is also striking: whereas *OPR* genes are randomly dispersed throughout the *Chlamydomonas* or *Volvox* genomes, almost all *NCL* genes are clustered on chromosome 15 or on an unplaced scaffold_19 which may later fill one of the sequence gaps in Chromosome_15. The rapid evolution of this subfamily likely rests on tandem duplications,

giving rise to seven sub-clusters of closely related genes. This tandem organisation would favour unequal crossing-overs, leading to gene duplication or loss and to repeat duplication. Interallelic recombination and gene conversion could also participate in the shuffling of repeats and thus to the diversification of binding sites, contributing to the neo-functionalisation of the newly formed paralogs and hence to their stabilisation. Indeed, we found that *NCL* genes evolve under diversifying selection pressure, with specific positions showing a high dN/dS ratio. The spontaneous appearance of the *ncc1* and *ncc2* mutations in laboratory conditions is consistent with this diversification still being an active process. *NCL* genes also can decay, as evidenced by the presence of inactivated and truncated *NCL* genes. Thanks to this vigorous "birth-diversification-death" process, the *NCL* family thus represents a constant source of RNA-binding proteins with new target specificities, and possibly of new sequence-specific endo-ribonucleases.

The properties of the *NCL* family are reminiscent of two other examples of rapidly evolving gene families in the plant kingdom. Restorer of Fertility-Like (RFL)-PPR proteins are distributed throughout higher plants and include Restorers of fertility (Rfs) proteins characterised in radish, petunia, or rice ((Fujii et al., 2011), reviewed in (Dahan and Mireau, 2013)). Rfs repress the expression of mitochondrion-encoded chimeric open reading frames that are generated by recombination between different copies of the mitochondrial genome and cause cytoplasmic male sterility (CMS) in various crop species. RFL proteins are considered as a reservoir for the evolution of Rfs counteracting the expression of new CMS genes. As *NCL* genes, most *RFL* genes are clustered -in two regions of chromosome 1 in *Arabidopsis thaliana*, or on chromosome 10 in rice-, and it is believed that unequal crossovers and local duplications within these clusters have led to the expansion of the RFL subfamily (Hernandez Mora et al., 2010). Their sequence evolves rapidly, with large divergence between species or even between different accessions (Jonietz et al., 2010). *RFL* genes are under diversifying selection (Geddy and Brown, 2007; Fujii et al., 2011), especially those residues of the PPR motifs that are involved in nucleotide recognition, which thus obviously favours the appearance of new RNA binding specificities (Geddy and Brown, 2007; Fujii et al., 2011; Barkan et al., 2012).

Similarly, pathogen resistance (*R*) genes, which activate plant defence reactions upon recognition of specific pathogen effectors, evolve rapidly in an "arms race" against the constantly evolving plant pathogens. *R* genes predominantly belong to the large and highly dynamic family of plant Nucleotide binding site Leucine-Rich repeat (NLR) proteins (reviewed in (DeYoung and Innes, 2006; Ye and Ting, 2008; Qi and Innes, 2013). They are

often clustered on the genome, a situation that, again, favours the expansion and evolution of the gene family (Michelmore and Meyers, 1998) and are under diversifying selection that targets mainly the solvent-exposed residues involved in protein-protein interaction (Wulff et al., 2009; Seeholzer et al., 2010). It is of note that the new alleles created by these diversifying selection processes in *RFL* and *NLR* families confer a dominant phenotype of fertility or pathogen resistance to their host organism (Elkonin, 2005; Song et al., 2006; Moffett, 2009). In many respects, *Chlamydomonas NCL* genes thus strikingly resemble the *RFL* and *NLR* genes families of higher plants.

The elusive physiological function of *NCL* proteins.

In the above examples, the evolutionary drive for the expansion of gene families are biotic stress and genome warfare, respectively. However, the selective pressure that led to the expansion of *NCL* genes in *Chlamydomonas*, after its separation from the *Volvox* lineage, is not known. To address this question, some knowledge of the physiological function of the wild-type NCC1 and NCC2 proteins would be of interest.

We fail to find additional phenotypes conferred by the *ncc1* or *ncc2* mutations. Either the gain of function provided by the substitutions within $NCC1^M$ and $NCC2^M$ preserves the original function of the wild-type proteins or these functions are subtle enough to have escaped our phenotypic analysis. A similar situation was reported in *Arabidopsis* where inactivation of the *RFL* genes *RPF1* or *RPF2*, while modifying the processing of a few mitochondrial transcripts (respectively *nda4* or *cox3* and *nad9*), does not alter the accumulation of their gene products, nor lead to any obvious phenotype (Jonietz et al., 2010; Holzle et al., 2011). Furthermore the *RPF2* gene seems inactivated in some accessions (Forner et al., 2008; Jonietz et al., 2010). It is also possible that, in contrast to their mutant allele, *NCC1* and *NCC2* have no function, would the gene duplication be too recent to have elicited the recruitment of their protein products for some physiological process. However, their expression seems regulated, at least at the RNA level: cursory examination of their expression profiles on the Phytozome browser suggests a transient repression of *NCC2* during nitrogen starvation (GSE34585) and a repression of *NCC1* by H_2O_2 treatment (GSE34826).

The spontaneous appearance of the *ncc1* and *ncc2* mutations likely illustrates a "trial and error" process that uses, as a template for genetic modification, genes whose physiological function is not essential or redundant. It may ultimately lead to the recruitment of a new RNA target, which would provide the protein with some functional significance if placed in appropriate environmental conditions. This is well illustrated by the experimental

conditions in which we recovered the *ncc2* strain. It shows a leaky cytochrome b_6f-defective phenotype that appeared when plating a non-phototrophic chimeric strain in phototrophic conditions (see Fig. 1). However it would be readily counter-selected if it occurred in cells with a wild-type chloroplast genome kept in phototrophic conditions.

Materials and methods

Strains and growth conditions

All Chlamydomonas strains, listed in Tables I and II, were grown in Tris-acetate-phosphate (TAP) medium, pH 7.2 (Harris, 1989), under continuous light (5-10 μE m^{-2} s^{-1}) at 25°C on a rotary shaker (150 rpm). Strains *ncc1* and *ncc1* {FAFA} were previously described as *mda1*-ncc1 and *mda1*-ncc1{FAFA} in (Drapier et al., 2002). Crosses, described in Table I, were performed according to Harris (1989). Vegetative diploids were selected on arginine-free plates from crosses between strains *ncc2 arg2* and *ncc2 arg7*, carrying the complementing, but not recombining, *arg2* and *arg7* mutations (Ebersold, 1967; Harris, 1989). After 12 days in low light, dark green colonies comprised of large cells were checked by PCR for the presence of both *mt* loci (Werner and Mergenhagen, 1998; Zamora et al., 2004).

Constructs and nucleic acids manipulations

Standard nucleic acid manipulations were performed according to (Sambrook et al., 1989). Primers used in that study are listed in Table SII.

5'*petA-psbB* chimera.

The *petA* 5'UTR and promoter regions were excised with *Eco*RV and *Nco*I from plasmid p5F (Choquet et al., 1998) and cloned into the p38A.NcoI vector (Vaistij et al., 2000) digested with *Stu*I and *Nco*I to yield plasmid p5'*petA-psbB*. The spectinomycin resistance cassette, excised from plasmid pUC-ATPX-AAD (Goldschmidt-Clermont, 1991) with *Sma*I and *Eco*RV, was inserted in the unique *Avr*II restriction site, blunted with Klenow enzyme, in reverse orientation with respect to the *psbB* gene, yielding plasmid pK5'*petA-psbB*, allowing the selection of transformants on spectinomycin-supplemented TAP plates.

pΔ*f*::α_{Tr}

In this plasmid, a truncated version of the *atpA* coding sequence was fused the *petA* lumen targeting peptide and replaced the sequence encoding the mature cytochrome *f* protein. The truncated *atpA* coding sequence was amplified in two fragments from plasmid p*atpA2* (Ketchner et al., 1995) using primers *atpA*$_{Fus}$FW and *atpA*$_{Fus}$RV1 on one hand and primers *atpA*FW and *atpA*$_{Fus}$RV2 on the other hand. The 254 bp fragment amplified with primers *atpA*$_{Fus}$FW and *atpA*$_{Fus}$RV1 was digested with *Hind*III and *Pst*I and cloned into vector p*f*::H_6 (Choquet et al., 2003) digested with the same enzymes to create plasmid pΔ*f*::α_{Tr}-int. The 965 bps fragment amplified with primers *atpA*FW and *atpA*$_{Fus}$RV2 was digested with *Eco*RI and

*Pst*I and cloned into vector p*f*::H_6-int, digested with the same enzymes to create plasmid pΔ*f*::α_{Tr}. The *aadA* spectinomycin resistance cassette was then inserted at the *Hinc*II site, upstream and in reverse orientation with respect to the *petA* coding sequence, to yield plasmids pKΔ*f*::α_{Tr}.

p5'*dAfR*

To create plasmid p5'*dAfR*, comprising the *petA* coding sequence expressed under the control of the *atpA* 5'UTR and the *rbcL* 3'UTR, we used plasmid pFADBE1 (Kuras and Wollman, 1994), which encompasses the *petA* genomic region, but where the *petA* gene was replaced by the *aadA* cassette. The *aadA* coding sequence was excised with *Nco*I and *Pst*I from pFADBE1 and replaced by the *petA* coding sequence, amplified from plasmid pWF (Kuras and Wollman, 1994) using primers AFRF_FW and AFRF_RV and digested by *Nco*I and *Pst*I, two restriction sites introduced in the sequence of the primers. Then, the 5'*psaA-aadA-3'rbcL* spectinomycin resistance cassette (Wostrikoff et al., 2004), excised with *Sma*I and *Eco*RV, was introduced at the *Hinc*II site upstream of the chimeric *petA* gene, in reverse orientation with respect to this latter.

Frameshifted petA gene: $pf_{31}St$ and $pf_{130}St$

To create frame shifts after the 31st and 130th codons of the *petA* gene, allowing respectively the synthesis of a total of 42 and 145 amino acids, plasmid pWF was digested by *Hind*III and *BstE*II, treated with Klenow and religated on itself to yield plasmids pf_{42}St and pf_{145}St. The 5'*psaA*-driven spectinomycin resistance cassette (Wostrikoff et al., 2004) was then inserted at the *Hinc*II site, upstream and in reverse orientation with respect to the *petA* coding sequence, to yield plasmids pKf_{42}St and pKf_{145}St.

Chimeric petD genes.

We generated the plasmid p*petD*, which contains a single *Swa*I site, within the *petD* 5'UTR, by digesting plasmid pWQ (Kuras and Wollman, 1994) with *Sac*I et *Afl*II; the resulting 5558 bp fragment was treated with the Klenow enzyme and religated on itself. To insert the target of the mutated version of NCC2 into the *petD* 5'UTR or within the *petD* coding sequence, a PCR fragment of 462 bps was amplified using primers Pet5::T2FW and petD$_{Cod}$::T2RV (see suppl. Table III for the list of the primers used in that study) and plasmid p*petD* as a template. This amplicon was either digested with *Swa*I and *Hind*III and the 88 bp fragment was cloned into vector p*petD* digested with the same enzymes to create plasmid p5'*petD*::T2 or with *Hind*III and *Pst*I and the 358 bps fragment was cloned into p*petD* digested by the same enzymes to create plasmid p*petD*$_{Cod}$::T2.

These two plasmids were then digested by *Avr*II, Klenow-treated, and ligated to a recycling 5'*psaA*-driven spectinomycin resistance cassette. Indeed, the *ncc1* mutation targeting the *atpA* mRNA, we avoided using the *5'atpA*-driven recycling cassette (Fischer et al., 1996) as a selectable marker and constructed a 5'*psaA*-*aadA*-3'*rbcL* recycling cassette. To this end, the *atpA* 5'UTR was excised from plasmid p*aadA*485 (Fischer et al., 1996) with *Nco*I and *Nru*I and replaced by the *psaA* 5'UTR, excised from plasmid p*faA*K (Wostrikoff et al., 2004) by *Eco*RV and *Nco*I, to create plasmid p5'aA-$aadA_{485}$. The cassette was excised from this vector by digestion with *Sac*I and *Kpn*I and Klenow treatment. In the final plasmids pK^{r}5'*petD*::T2 and pK^r*petD*$_{Cod}$::T2, the *aadA* cassette was transcribed in reverse orientation with respect to the *petD* gene.

Mutation of the NCC1^M target

We mutated the target of the mutated version of the NCC1 protein by two-steps megaprime PCR (Higuchi, 1990): two pairs of primers, *atpA*$_{Ext}$FW/*atpA*MT1RV and *atpA*MT1FW/*atpA*$_{Ext}$RV, allowed the amplification from plasmid p*atpA2* of two partially overlapping fragments that were mixed and used as templates in a third PCR with the external primers *atpA*$_{Ext}$FW and *atpA*$_{Ext}$RV. The final amplicon was digested by *Mfe*II and *Pac*I, two restriction sites on either sides of the mutation, and cloned into plasmid pK^r*atpA*$_{300}$St (Drapier et al., 2007) digested with the same enzymes to create plasmid pK^r*atpA*$_{Mut}$T2.

Mutation of the NCC2^M target

We destroyed the target of NCC2^M by the same two steps PCR procedure using the mutagenic primers *petA*MT2FW and *petA*MT2RV and the external primers *petA*$_{Ext}$FW and *petA*$_{Ext}$RV to amplify from plasmid template pWF a 1053 bp fragment. This amplicon was digested by *Bgl*II and *Acc*I, two restriction sites on either sides of the mutation, and cloned into plasmid pKWF$_{Stop}$ (Boulouis et al., 2011) digested with the same enzymes to create plasmid pK*petA*$_{Mut}$T2.

Transformation vector for expression of the mutated NCC1 and NCC2 proteins.

Because of the high percentage of similarity between the paralogous genes of the *NCL* cluster, designing specific primers to amplify the mutant *ncc1* and *ncc2* genes turned out to be difficult. We thus ordered the synthetic DNA sequences indicated in the Supplemental Material and Methods section (**Genscript**, Piscataway, NJ). They were digested by *Eco*RI and *Bgl*II (*ncc1*-HA) or *Eco*RI and *Bam*HI (*ncc2*-HA) and cloned into the vector pJHL (kindly provided by Dr Jae-Hyeok Lee, University of British Columbia) digested by *Eco*RI and *Bam*HI.

All DNA constructs were sequenced before transformation in *C. reinhardtii*. Northern blot analyses were carried out as described in (Drapier et al, 2002), using ^{33}P-labelled probes described in (Eberhard et al., 2002). Transcript accumulation was quantified from Phosphor-Imager scans of the blots, as described in (Choquet et al., 2003).

Chloroplast translation was arrested by supplementing cells grown in TAP medium (2 x 10^6 cells mL^{-1}) with lincomycin (final concentration 500 µg mL^{-1}) at t = 0. Aliquots were taken at the indicated time points and briefly chilled on ice before RNA extraction.

Map-based localisation of the *ncc2* mutation.

To localise the *ncc2* mutation, a first set of about fifty *ncc2* progeny (based on their fluorescence phenotype) was selected out of independent meiosis from the cross *ncc2* x S1-D2. After Chelex-based DNA extraction (Neubauer et al., 2009), AFLP markers (Kathir et al., 2003; Rymarquis et al., 2005) allowed to determine the proportion of each parental version of the marker on each chromosome arm by PCR. Once linkage (96 %) to the ZYS3 marker established, new AFLP markers, based on identified polymorphisms in S1-D2 ESTs or on putative differences in tandem repeat copy numbers, were designed along chromosome 15 (Suppl. Table EII) to restrain the region containing the *ncc2* mutation. To observe rare crossing-over events, this analysis was performed on 500 independent meioses and allowed to restrict the location of the mutation to a 405 kbp region.

Genomic DNA preparation, whole genome sequencing and data analysis.

DNA from wild-type and {FAFA} *ncc1 ncc2* double mutant strains was extracted with the DNAeasy® Plant Maxi Kit (Qiagen), according to the manufacturer's protocol, starting from 100 mL of stationary cultures.

Genome sequencing was performed using the high throughput, short read, Illumina technology. Sequencing was done at the Tufts University Core Facility, Boston, USA, on a HiSeq 2000 instrument in paired-end mode, 2 x 100 nt. Libraries had insert sizes of ~300 bp. About 180 million read pairs were generated. Reads were mapped simultaneously onto the nuclear, chloroplast, and mitochondrial genomes of *C. reinhardtii* using BWA 0.6.0-r85 (Li and Durbin, 2009). The organelle genomes were taken from GenBank (accessions FJ423446, strain CC-503 cw92 *mt*+, and CRU03843, strain CW-15-2), and version 5 of the nuclear genome was used (strain CC-503 cw92 *mt*+; obtained from Phytozome). The BWA "aln" and "sampe" commands were run with default parameter values, except for the following options: "-q -1 -R 10 -o 2 -e 0 -l 30 -t 20" for "aln", and "-a 600 -n 9999 -N 9999" for "sampe". The

alignment files created by BWA in SAM format were then converted to BAM format, indexed, and duplicate read pairs were removed using samtools 0.1.16 (Li et al., 2009). Read mapping data were visualised in IGV 2.0 (Robinson et al., 2011; Thorvaldsdottir et al., 2013).

The duplicate-filtered alignment files were then fed to the SVMerge 1.1r32 pipeline (Wong et al., 2010) to detect single nucleotide polymorphisms (SNPs), short insertions and deletions (indels), and large structural variants (SVs). SVMerge included the following tools: breakdancer 1.1 (Chen et al., 2009), pindel 0.2.4q (Ye et al., 2009), cnD 1.3 (Simpson et al., 2010), and SECluster (bundled with SVMerge). The results from InGAP-sv 2.8.1 (Qi et al., 2010; Qi and Zhao, 2011) were also integrated in the SVMerge pipeline. Breakdancer was run with the following options: "-q 35 -c 3 -r 3 -y 40 -m 10000000"; pindel was run with the following options: "-T 1 -x 9 -e 0.02 -u 0.05 -a 1 -m 3 -n 50 -v 50 -d 30 -A 35 -M 6"; cnD was run with the following options: "--threshold=0.5 --window=5"; SECluster was run with the following options: "-q 35 -m 6 -c 6 -r {1} -x 10000"; inGAP-sv was run with default parameters, but a maximum coverage of 1000. In the SVMerge configuration file, the following cut-off scores were set: "BDscore=40", "BDrs=3", and "PDscore=200". Velvet 1.1.05 (Zerbino and Birney, 2008) and exonerate 2.2.0 (Slater and Birney, 2005) were run for local assemblies of SVs in SVMerge. The following Velvet parameters were specifically set: "hashlen=25, exp_cov=auto, cov_cutoff=3". SVMerge requires the BEDtools package (Quinlan and Hall, 2010); version 2.16.2 was used.

Based on version 5.0 of the gene predictions (obtained from Phytozome), SNPs and indels were annotated using SHOREmap_annotate from the SHOREmap 1.2 package (Schneeberger et al., 2009) to determine in which features they were located (gene, UTR, CDS, intron), and whether they would cause synonymous or non-synonymous changes.

The phylogenic tree was constructed on the Phylogeny.fr platform (Dereeper et al., 2008), including the following steps: sequences were aligned with MUSCLE (v3.7) (Edgar, 2004) configured for highest accuracy (MUSCLE with default settings), cured with Gblocks (v0.91b), using relaxed parameters. The phylogenetic tree was reconstructed using the maximum likelihood method implemented in the PhyML program (v3.0) (Guindon et al., 2010) with reliability for internal branch assessed using the aLRT test (SH-Like) (Anisimova and Gascuel, 2006) and visualised using TreeDyn (v198.3) (Chevenet et al., 2006).

Bioinformatic analysis of positive (diversifying) selection

Multiple sequence alignment of the NCL proteins was generated with ClustalW (Sievers et al., 2011) and manually refined to optimize conservation of the repeats. Pal2NAL

v.14 was used to align the CDS based on this alignment (Suyama et al., 2006), and MEGA to generate the phylogenetic tree using maximum likelihood (Tamura et al., 2013). Diversifying selection was analysed using PAML v4.7a (routine codml) as described in (Yang, 2007), using the graphic interface in PAMLX (Xu and Yang, 2013). The likelihood of neutral selection models (M1 and M7) was compared with that of models allowing a class with $\omega>1$ (M2 and M8). Model 1 allows two classes of codon with $\omega=0$ (negative selection) and $\omega=1$ (neutral selection), while model 2 in addition allows a third class with $\omega>1$ (positive selection). Models 7 and 8 allow a continuous β-distribution of ω values <1, with an additional class with $\omega>1$ in model 8. (Yang and Swanson, 2002). Posterior probability of positive selection at each site was calculated the using Bayes-Empirical Bayes method (Anisimova et al., 2002).

Transformation experiments

Cells were transformed by tungsten particle bombardment (Boynton et al., 1988) as described in (Kuras and Wollman, 1994). Transformants, listed in Table III, were selected on TAP-Spec (100 $mg.ml^{-1}$) and subcloned on TAP-Spec until they reached homoplasmy, assessed by RFLP analysis. At least three independent transformants were analysed for each transformation. Phenotypic variations between independent transformants proved negligible.

Nuclear transformation of the wild-type was performed by electroporation, as described in Raynaud *et al.*, 2007, with the following parameters: 25 µF, 1000 V cm^{-1}. Transformants were selected on paromomycin (5 $\mu g.ml^{-1}$) supplemented plates.

Protein analysis

Pulse-labelling experiments, protein electrophoresis and immunoblots were performed on exponentially growing cells (2 x 10^6 $Cells.ml^{-1}$) according to (Kuras and Wollman, 1994). Cell extracts, loaded on equal chlorophyll basis, were analysed by SDS–PAGE (12– 18% acrylamide, 8M urea). Anti-OEE2, and -cytochrome *f* antibodies, used for [^{135}I]protein A detection were raised in the laboratory against proteins isolated from *C. reinhardtii* and are respectively described in (de Vitry et al., 1989; Kuras and Wollman, 1994). HA-tagged $NCC1^M$ and $NCC2^M$ proteins were detected by ECL using the monoclonal antibody anti HA.11 (Covance), and horseradish peroxidase–conjugated antibody against mouse IgG (Promega).

Fluorescence measurements

were performed on dark-adapted liquid cultures using a home built spectrofluorimeter according to (Zito et al., 1997).

Acknowledgments:

We thank Dr Jae-Hyeok Lee (University of British Columbia) for his kind gift of the pJHL *Chlamydomonas* transformation vector, D. B. Stern for providing markers for molecular mapping prior to publication (Rymarquis et al., 2005) and all members of UMR7141 for stimulating discussions and/or critical reading of the manuscript. This work was supported by Unité Mixte de Recherche 7141, CNRS and Université Pierre et Marie Curie, Paris 06, by the European Community ("SunBioPath" Contract FP7-KBBE-2009-3-02; GIAVAP Contract FP7-KBBE.2010.3. GA No. 266401), by Agence Nationale de la Recherche (ANR) contract (ChloroMitoCES: BLAN-NT09_451610 and ChloroRNP: ANR-13-BSV7-0001-01) and by the "Initiative d'Excellence" program (Grant "DYNAMO," ANR-11-LABX-0011-01). A.B. was "Attachée Temporaire d'Enseignement et de Recherche" at Université Pierre et Marie Curie. N. T. was supported by GIAVAP and Dynamo.

Author contributions:

AB, DD, FAW and YC designed research; AB, DD, HR, KW, GBJ and YC performed research; NT KP and OV contributed to analytic tools and bioinformatics analysis; AB, DD, FAW and YC analysed data; and AB, DD, FAW and YC wrote the paper.

Conflict of interest:

The authors declare that they have no conflict of interest

Figures legend: 8825

Figure 1: Isolation of the *ncc2* mutant.

A) Schematic description of the *psbB* gene in wild type and {5'*cA*CP47} transformed cells.

Bent arrows indicate promoters, while the insertion site and the orientation of the spectinomycin resistance cassette, symbolised by K, are indicated.

B) Fluorescence induction kinetics of dark adapted wild-type and {5'*cA*CP47} cells.

The nearly constant fluorescence intensity over time in strain {5'*cA*CP47}, as well as its high initial level, almost similar to the stationary level, are typical of leaky PSII mutants.

C) Phenotypic characterisation of {5'*cA*CP47} transformed strains.

(Left) CP47 accumulation probed with a specific antibody in two independent transformants, in a dilution series of the wild type and in the *mbb1*-222E strain, defective for the accumulation of the *psbB* mRNA (Monod et al., 1992), as a control for the specificity of the antibody. OEE2, whose accumulation is independent of that of the core PSII complex (de Vitry et al., 1989), serves as a loading control. (Right) *psbB* mRNA accumulation in the same strains. Because of the larger size of the *petA* 5'UTR, compared to the *psbB* 5'UTR, the chimeric mRNA migrates more slowly than the endogenous *psbB* transcript upon electrophoresis.

D) Growth properties of the wild-type, {5'*cA*CP47} and Su0-T9 strains.

Drops of liquid cultures (2 x 10^6 cells.mL^{-1}) were spotted on Tap medium and grown under dim light (10 µE.m^{-2}.s^{-1}, top), on Minimum medium under high light (150 µE.m^{-2}.s^{-1}, intermediate panel) or on Tap supplemented with spectinomycin (500 µg.mL^{-1}). Pictures were taken after 10 days of growth.

E) Phenotypic characterisation of the Su0-T9 strain.

Accumulation of CP47 and cytochrome *f* (Left) and of the *psbB* mRNA (Right) in the same strains. In RNA blot experiments the *mbb1*-222E strains provides a control for the specificity of the probe and *psbA* was used as a loading control.

Figure 2: The dominant *ncc2* mutation confers a b_6f leaky phenotype due to a reduced accumulation of the *petA* mRNA.

A: Fluorescence induction kinetics of dark-adapted *ncc2* and wild-type cells. Black and grey curves respectively show the kinetics recorded in the absence and in the presence of

DCMU (5 μM) that blocks electron transfer downstream of photosystem II. Maximal fluorescence levels in the presence of DCMU were normalised to 1.

B: Cytochrome *f* accumulation in wild-type and *ncc2* strains, probed with a specific antibody. OEE2 provides a loading control.

C: Translation of chloroplast genes determined by 5 min ^{14}C-acetate (5 μCi.ml^{-1}) pulse-labelling experiments performed in the presence of cycloheximide (10 μg.ml^{-1}) to block cytosolic translation. The arrow indicates the position of cytochrome *f*.

D: Accumulation of representative transcripts for chloroplast photosynthesis genes in wild-type and *ncc2* strains, assessed by RNA blots using the probes indicated on the right of the panel. For the *petA* gene, the diamond indicates the mature transcript, while asterisks point to precursor RNA species.

E: Accumulation of *petA* mRNA and cytochrome *f*, determined as above, in diploid strains either homozygous or heterozygous for the *ncc2* mutation. The wild-type and *ncc2* strains are shown for comparison. *psbB* mRNA and OEE2 respectively provide loading controls in RNA- and immuno-blots.

Figure 3: The *ncc2* mutation targets the *petA* coding sequence.

A: Scheme of the chimeras used in this figure. The sequence encoding the mature cytochrome *f* is shown as dark grey rectangle and that encoding the lumen targeting peptide as hatched rectangle. *petA* 5' and 3' UTRs are represented by thinner light grey rectangles. *petA* promoter is indicated by a bent arrow. In Δ*f*::α_{Tr} chimera, the region encoding mature cytochrome *f* was replaced by the first 944 nucleotides of the *atpA* coding sequence, depicted as a blue rectangle. In *dAfR* chimera the *petA* promoter and 5'UTR regions were replaced by the corresponding *atpA* regions (pale blue rectangle), while *petA* 3'UTR was replaced by that of the *rbcL* gene (brown rectangle).

B: Transcript accumulation in tetrad progeny from the cross {Δ*f*::α_{Tr}} x *ncc2* and in wild-type and parental strains. RNA blots were hybridised with probes derived from *petA*, *atpA* and *psaA* (loading control) coding sequences, as indicated on the left. Positions of the endogenous mono- and di-cistronic *atpA* transcripts are indicated by ● and ■, respectively. Arrow points to the position of the major Δ*f*::α_{Tr} chimeric transcript, while * indicates a minor cotranscript that includes the downstream *aadA* resistance cassette.

C: *petA* and *atpA* (loading control) transcript accumulation in tetrad progeny from the cross {*dAfR*} x *ncc2* and in wild-type and parental strains. Arrow points to the position the chimeric *dAfR* transcript.

Figure 4: The *ncc2*, but not *ncc1*, phenotype is observed only when the target RNA is translated.

A: *atpA* transcript accumulation in tetrad progeny from the cross {$atpA_{St}$} x *ncc2* and in wild-type and parental strains. Position of the four transcripts from the *atpA* tetracistronic transcription unit is indicated. *rbcL* mRNA: loading control.

B: *petA* transcript accumulation in tetrad progeny from the cross {$petA_{St}$} x *ncc2* and in wild-type and parental strains (loading control: *atpA*).

C and D: Accumulation of *atpA* (C) and *petA* (D) transcripts in *ncc1* (C), *ncc2* (D) and wild-type strains incubated in the presence of lincomycin for the indicated times. Loading controls: *psbD* (C), *atpA* (D).

Figure 5: Narrowing down the target sequences of the *ncc1* and *ncc2* mutations.

A: Schematic representation of the mutant *petA* genes, presented as in Fig. 3A. F points to the position of the introduced frameshifts, while grey and white rectangles respectively show translated and untranslated *petA* sequences.

B & C: *petA* transcript accumulation in tetrad progeny from crosses {f_{145}St} x *ncc2* (B) and {f_{42}St} x *ncc2* (C). *atpA* (C) and *psbD* (B) mRNAs serve as loading controls.

D: Accumulation *of atpA*-hybridising transcripts in wild-type and *ncc1* strains transformed with the $\Delta f::\alpha_{Tr}$ chimera (Fig. 3A). The probe hybridises with the chimeric transcript, either alone ($\Delta f::\alpha_{Tr}$) or cotranscribed with the *aadA* cassette (*) and with endogenous mono- (●) and di-(■)-cistronic *atpA* mRNA.

Figure 6: The *NCC1* and *NCC2* genes.

A: (Top) Genetic and molecular map of the *ncc1* and *ncc2* loci. Locations on chromosome 15 of the *ZYS3* marker and *ncc1* and *ncc2* mutations are shown, along with genetic distances. The origin of the discrepancy between genetic distances determined in the three point test has not been investigated but may result from the poor fluorescence identification of some double mutants. The pink rectangle represents the molecular region containing the *ncc2* mutation, as determined by map-based cloning. (Bottom) Physical map of

the *NCL* gene cluster on chromosome 15: *NCL* genes are drawn in red, with the exception of *NCC1* and *NCC2* drawn in blue. Non OPR genes are drawn in grey.

B: Schematic representation of the NCC1 and NCC2 protein. White rectangles depict the chloroplast transit peptide predicted by the ChloroP program. Dark grey boxes represent the OPR repeats, while lighter grey boxes indicate additional, more degenerated, repeats. Punctuated rectangles show the position of the RAP domains. The highly conserved region between the two proteins is indicated. Also shown are positions of the two substitutions in the *ncc1* and *ncc2* strains (written in red).

C: Alignment of OPR repeats in the NCC1 (top) and NCC2 (bottom) proteins, with residues corresponding to the consensus (bottom) shaded in grey. Mutated amino acids in *ncc1* and *ncc2* strains are written in red.

D) Phylogeny of NCL proteins. Sequence distance Maximum Likelihood tree of the NCL proteins, using Chlre_OPR68 as an out-group. Branch length represents the estimated rate of amino acid substitution. Coloured boxes remind of the genomic location of the corresponding NCL genes, as indicated in the bottom panel of Fig. 6A. NCC1 and NCC2 are written in blue.

Figure 7: Allotopic expression of a mutated version of the NCC1 (/NCC2) protein confers the *ncc1* (/*ncc2*) phenotype to the transformed strains.

A: Accumulation of the *petA* and *atpA* mRNAs in wild-type, *ncc1*and *ncc2* strains and in two transformed strains expressing respectively $NCC1^{M}$ and $NCC2^{M}$, as shown by immunodetection using an antibody against their HA tagged version (upper panel).

B: (left) Increasing accumulation of *atpA* mRNA in a series of transformants accumulating decreasing amounts of $NCC1^{M}$. (right) Two transformants illustrating the negative correlation between accumulations of *petA* mRNA and of $NCC2^{M}$.

Figure 8: Identification of the target of $NCC1^{M}$

A: Location of the target of $NCC1^{M}$, written in blue along the *atpA* gene. The mutation introduced in the *atpA*$T1^{M}$ construct is shown in red.

B: Accumulation of the *atpA* transcript in wild-type and *ncc1* strains transformed with the *atpA*$T1^{M}$ construct. Independent transformants are presented for each background. Untransformed wild-type and *ncc1* strains are shown for comparison (*psbD*: loading control).

Figure 9: Identification of the target of $NCC2^{M}$

A: Location of the target of NCC2^{M}, written in blue along the *petA* gene. Silent mutations introduced in the *petA*T2^{M} construct are written in red. Residual translation of *petA* in the f_{42}St mutant, downstream of the frameshift, is shown. The *Hind*III site used to introduce the frameshift indicated in red is underlined. Boxed nucleotides are those which should be covered by ribosome stalling on the premature Stop codon.

B: Accumulation of the *petA* transcript in wild-type and *ncc2* strains transformed with the *petA*T2^{M} construct. Three independent transformants (#1-#3) are presented for each background. Untransformed wild-type and *ncc2* strains are shown for comparison (loading control: *psbD*).

C: Insertion sites of the NCC2^{M} target within the *petD* gene. Schematic representation of the *petD* gene with the 5'UTR and coding sequence respectively drawn as thin light grey and thick dark grey rectangles, while the 3 white boxes represent transmembrane helices. Relevant restriction sites (*Swa*I, *Hind*III and *Pst*I) are indicated. Nucleotide regions surrounding the NCC2^{M} target are shown with relevant restriction sites underlined. Note that the target of NCC2^{M} is translated in a different frame in the mutated *petD* gene and in *petA* (lower line).

D: Accumulation of *petD* transcript in wild-type and *ncc2* strains transformed with the 5'*petD*::T2 (left) and *petD*$_{Cod}$::T2 (right) constructs. Three independent transformants (#1-#3) are presented for each background. Untransformed wild-type and *ncc2* strains are shown for comparison. A *petA* probes reveal *ncc2* background (loading control: *psbD*).

E: *petD* and *petA* mRNAs accumulation in the *ncc2* {*petD*$_{Cod}$::T2} transformant #3, either untreated or treated with lincomycin for 4 h. Untreated wild-type and similarly treated *ncc2* strain are shown for comparison. *atpA* and *psbD* provide loading controls.

Table I: Strains generated by crosses during this study.

***mt*⁻ parent**	***mt*⁺ parent**	**progeny**
ncc2 {5cACP47} (Su0-T9)	wild-type	*ncc2*
ncc2	*arg2* [1]	*ncc2 arg2*
ncc2	*arg7* [1]	*ncc2 arg7*
ncc2 arg2	*ncc2 arg7*	Diploid *ncc2/ncc2*
ncc2 arg2	*arg7*	Diploid *ncc2/NCC2*
arg2	*arg7*	Diploid *NCC2/NCC2*
ncc2	{Δ*f*::α_{Tr}} [2]	*ncc2* {Δ*f*::α_{Tr}}
ncc2	{5'*dAfR*}[2]	*ncc2* {5'*dAfR*}
ncc2	{*petA*$_{St}$} [3]	*ncc2* {*petA*$_{St}$}
ncc2	{*f*$_{42}$St}[2]	*ncc2* {*f*$_{42}$St}
ncc2	{*f*$_{145}$St}[2]	*ncc2* {*f*$_{145}$St}
ncc2	{FAFA} *ncc1*[4]	*ncc1 ncc2* {FAFA}
ncc1	{*atpA*$_{St}$}[5]	*ncc1* {*atpA*$_{St}$}

References: [1] (Ebersold, 1967; Harris, 1989); [2]: this work; [3] (Boulouis et al., 2011); [4] (Drapier et al., 2002); [5] (Eberhard et al., 2011).
By convention, chloroplast genotypes, when relevant, follow the nuclear genotype and are written between braces.

Table II: Strains generated by transformation during this study.

	Recipient Strain	Transforming Plasmid	Transformed Strain
Chloroplast transformation[a]	wild-type *mt*$^+$	pK5'*petA-psbB*	{*5cACP47*}
	wild-type *mt*$^+$	pKΔf::α_{Tr}	{Δf::α_{Tr}}
	wild-type *mt*$^+$	p5'*dAfRK*	{5'*dAfR*}
	wild-type *mt*$^+$	pf_{42}St	{f_{42}St}
	wild-type *mt*$^+$	pf_{145}St	{f_{145}St}
	ncc1 mt$^+$	pK*atpA*$_{St}$ [1]	*ncc1* {*atpA*$_{St}$}
	wild-type *mt*$^+$	pK*atpA*T1^M	{*atpA*T1^M}
	ncc1 mt$^+$	pK*atpA*T1^M	*ncc1* {*atpA*T1^M}
	wild-type *mt*$^+$	pK*petA*T2^M	{*petA*T2^M}
	ncc2 mt$^+$	pK*petA*T2^M	*ncc2* {*petA*T2^M}
	wild-type *mt*$^+$	pK5'*petD*::T2	{5'*petD::T2*}
	ncc1 mt$^+$	pK5'*petD*::T2	*ncc1* {5'*petD::T2*}
	ncc2 mt$^+$	pK5'*petD::T2*	*ncc1* {5'*petD*::T2}
	wild-type *mt*$^+$	pK*petD*$_{Cod}$::T2	{*petD*$_{CDS}$::T2}
	ncc2 mt$^+$	pK*petD*$_{Cod}$::T2	*ncc2* {*petD*$_{CDS}$::T2}
Nuclear Transformation[b]	wild-type	p*NCC1*^{M}-HA	*NCC1*^{M}*-HA*
	wild-type	p*NCC2*^{M}-HA	*NCC2*^{M}*-HA*

[a]: All recipient strains were *mt*$^+$ and spectinomycin sensitive: transformants were selected for resistance to spectinomycin (100 mg·mL^{-1}) under low light (5 µE.m^{-2}.s^{-1}) and subcloned in darkness until they reached homoplasmy.
b: Transformed strains were selected for resistance to paromomycin (10 mg·mL^{-1}) under low light (5 µE.m^{-2}.s^{-1}).
Reference: [1] (Eberhard et al., 2011)

Table III: Comparison of codon evolution models in *NCL* genes.

	Likelihood[1]		$2\Delta l$[2]	p-value[3]	p[4]	ω[5]
	Nearly neutral	positive selection				
A)	M1	M2	(M2 vs M1)			
	-43936,80	-43819,78	-234,04	$2,28\ E^{-102}$	0,074	2,88
B)	M7	M8	(M7 vs M8)			
	-43923,61	-43771,32	-304,57	$5,32\ E^{-133}$	0,153	2,05

A) Comparison of codon substitution models M1 vs. M2 in paralogous *NCL* genes

Model 1 allows two classes of codons (ω=0 -negative selection- and ω=1 -neutral selection-), while model 2 allows an additional class under positive selection (ω>1).

B) Comparison of codon substitution models M7 vs. M8 in paralogous *NCL* genes

Model 7 allows a continuous β-distribution of 0<ω<1, with an additional class (ω>1) in model 8.

[1] Log likelihood values are indicated for models allowing (models 2 and 8) or not (models 1 and 7) a class with ω>1.

[2] Likelihood ratio between the two models

[3] evaluated from χ^2 distribution (df=1).

[4] proportion of sites under positive selection (ω>1).

[5] mean ω value for sites under positive selection.

Supplemental information avalaible online:

- Fig. S1 Phenotype of the double mutant *ncc1 ncc2* {FAFA}, descendant of the cross *ncc1 mt*$^{+}$ {FAFA} x *ncc2 mt*$^{-}$, whose genome was sequenced.

- Fig. S2: alignment of NCL proteins.

The RAP domain may show endonucleolytic activity

- Fig. S3: I-TASSER alignment used for threading of the NCC1 RAP domain.
- Fig. S4: Comparison of NCC1 model with known structures of endonucleases

- Table SI: Genetic independence of the ncc2 mutation from MCA1 and TCA1 genes

- Table SII: markers designed to map the *ncc2* mutation.

- Table SIII: oligonucleotides used in this study.

- Table SIV: BLAST search-based identification of NCL proteins in Phytozome v5.5.

- Table SV: description of NCL proteins.
  Improved gene models for *NCL7, 8, 21, 30, 35*
  Intracellular targeting of NCL proteins.

- Supplemental Material and Method section.
  Sequence of the synthetic *NCC1*^{M} and *NCC2*^{M} genes.

- Bibliography for Supplemental Information.

Accession numbers: Sequence data for *NCC1*, *NCC2* and *NCL* genes from this article can be found in the Phytozome database, as indicated in Table EV. Other sequence data used in this article can be found in the GenBank/EMBL databases under the following accession numbers: *petA* (cytochrome *f*), *petD*: X72919.1; *psaB*: X05848.1; *psaC*: U43964.1; *psbA*, CAA25670; *psbB* (CP47): X64066.1; *psbD*: X04147.1; *OEE2*, M15187.1; *atpA*: X60298.1; *rbcL*: J01399.1.

References

Anisimova, M., and Gascuel, O. (2006). Approximate likelihood-ratio test for branches: A fast, accurate, and powerful alternative. Systematic biology **55,** 539-552.

Anisimova, M., Bielawski, J.P., and Yang, Z. (2002). Accuracy and power of bayes prediction of amino acid sites under positive selection. Mol Biol Evol **19,** 950-958.

Barkan, A. (2011). Expression of plastid genes: organelle-specific elaborations on a prokaryotic scaffold. Plant Physiol **155,** 1520-1532.

Barkan, A., and Goldschmidt-Clermont, M. (2000). Participation of nuclear gene in chloroplast gene expression. Biochimie **82,** 559-572.

Barkan, A., and Small, I. (2014). Pentatricopeptide repeat proteins in plants. Annu Rev Plant Biol **65,** 415-442.

Barkan, A., Rojas, M., Fujii, S., Yap, A., Chong, Y.S., Bond, C.S., and Small, I. (2012). A combinatorial amino acid code for RNA recognition by pentatricopeptide repeat proteins. PLoS Genet **8,** e1002910.

Blaby, I.K., Blaby-Haas, C.E., Tourasse, N., Hom, E.F., Lopez, D., Aksoy, M., Grossman, A., Umen, J., Dutcher, S., Porter, M., King, S., Witman, G.B., Stanke, M., Harris, E.H., Goodstein, D., Grimwood, J., Schmutz, J., Vallon, O., Merchant, S.S., and Prochnik, S. (2014). The *Chlamydomonas* genome project: a decade on. Trends Plant Sci **19,** 672-680.

Boch, J., Scholze, H., Schornack, S., Landgraf, A., Hahn, S., Kay, S., Lahaye, T., Nickstadt, A., and Bonas, U. (2009). Breaking the code of DNA binding specificity of TAL-type III effectors. Science **326,** 1509-1512.

Bogdanove, A.J., and Voytas, D.F. (2011). TAL effectors: customizable proteins for DNA targeting. Science **333,** 1843-1846.

Boulouis, A., Raynaud, C., Bujaldon, S., Aznar, A., Wollman, F.A., and Choquet, Y. (2011). The nucleus-encoded trans-acting factor MCA1 plays a critical role in the regulation of cytochrome *f* synthesis in *Chlamydomonas chloroplasts*. Plant Cell **23,** 333-349.

Boynton, J.E., Gillham, N.W., Harris, E.H., Hosler, J.P., Johnson, A.M., Jones, A.R., Randolph-Anderson, B.L., Robertson, D., Klein, T.M., and Shark, K.B. (1988). Chloroplast transformation in *Chlamydomonas* with high velocity microprojectiles. Science **240,** 1534-1538.

Chen, K., Wallis, J.W., McLellan, M.D., Larson, D.E., Kalicki, J.M., Pohl, C.S., McGrath, S.D., Wendl, M.C., Zhang, Q., Locke, D.P., Shi, X., Fulton, R.S., Ley, T.J., Wilson, R.K., Ding, L., and Mardis, E.R. (2009). BreakDancer: an algorithm for high-resolution mapping of genomic structural variation. Nature methods **6,** 677-681.

Cheong, C.G., and Hall, T.M. (2006). Engineering RNA sequence specificity of Pumilio repeats. Proc Natl Acad Sci U S A **103,** 13635-13639.

Chevenet, F., Brun, C., Banuls, A.L., Jacq, B., and Christen, R. (2006). TreeDyn: towards dynamic graphics and annotations for analyses of trees. BMC Bioinformatics **7,** 439.

Choquet, Y., and Wollman, F.A. (2002). Translational regulations as specific traits of chloroplast gene expression. FEBS Lett **529,** 39-42.

Choquet, Y., and Wollman, F.-A. (2008). The CES process. In *Chlamydomonas* Source Book, E.E. Harris, D.B. Stern, and W. g., eds (New York: Academic Press, Elsevier.

Choquet, Y., Zito, F., Wostrikoff, K., and Wollman, F.A. (2003). Cytochrome *f* Translation in *Chlamydomonas* Chloroplast Is Autoregulated by its Carboxyl-Terminal Domain. Plant Cell **15,** 1443-1454.

Choquet, Y., Stern, D.B., Wostrikoff, K., Kuras, R., Girard-Bascou, J., and Wollman, F.A. (1998). Translation of cytochrome *f* is autoregulated through the 5' untranslated region of *petA* mRNA in *Chlamydomonas* chloroplasts. Proc.Natl.Acad.Sci.U.S.A. **95,** 4380-4385.

Christian, M., Cermak, T., Doyle, E.L., Schmidt, C., Zhang, F., Hummel, A., Bogdanove, A.J., and Voytas, D.F. (2010). Targeting DNA double-strand breaks with TAL effector nucleases. Genetics **186,** 757-761.

Cooke, A., Prigge, A., Opperman, L., and Wickens, M. (2011). Targeted translational regulation using the PUF protein family scaffold. Proc Natl Acad Sci U S A **108,** 15870-15875.

Dahan, J., and Mireau, H. (2013). The Rf and Rf-like PPR in higher plants, a fast-evolving subclass of PPR genes. RNA biology **10,** 1469-1476.

de Vitry, C., Olive, J., Drapier, D., Recouvreur, M., and Wollman, F.A. (1989). Posttranslational events leading to the assembly of photosystem II protein complex: a study using photosynthesis mutants from *Chlamydomonas reinhardtii*. J Cell Biol **109,** 991-1006.

Deng, D., Yan, C., Pan, X., Mahfouz, M., Wang, J., Zhu, J.K., Shi, Y., and Yan, N. (2012). Structural basis for sequence-specific recognition of DNA by TAL effectors. Science **335,** 720-723.

Dereeper, A., Guignon, V., Blanc, G., Audic, S., Buffet, S., Chevenet, F., Dufayard, J.F., Guindon, S., Lefort, V., Lescot, M., Claverie, J.M., and Gascuel, O. (2008). Phylogeny.fr: robust phylogenetic analysis for the non-specialist. Nucleic Acids Res **36,** W465-469.

DeYoung, B.J., and Innes, R.W. (2006). Plant NBS-LRR proteins in pathogen sensing and host defense. Nature immunology **7,** 1243-1249.

Drapier, D., Girard-Bascou, J., Stern, D.B., and Wollman, F.A. (2002). A dominant nuclear mutation in *Chlamydomonas* identifies a factor controlling chloroplast mRNA stability by acting on the coding region of the *atpA* transcript. Plant J **31,** 687-697.

Drapier, D., Rimbault, B., Vallon, O., Wollman, F.A., and Choquet, Y. (2007). Intertwined translational regulations set uneven stoichiometry of chloroplast ATP synthase subunits. Embo J **26,** 3581-3591.

Drapier, D., Suzuki, H., Levy, H., Rimbault, B., Kindle, K.L., Stern, D.B., and Wollman, F.A. (1998). The chloroplast *atpA* gene cluster in *Chlamydomonas reinhardtii*. Functional analysis of a polycistronic transcription unit. Plant Physiol **117,** 629-641.

Eberhard, S., Drapier, D., and Wollman, F.A. (2002). Searching limiting steps in the expression of chloroplast-encoded proteins: relations between gene copy number, transcription, transcript abundance and translation rate in the chloroplast of *Chlamydomonas reinhardtii*. Plant J **31,** 149-160.

Eberhard, S., Loiselay, C., Drapier, D., Bujaldon, S., Girard-Bascou, J., Kuras, R., Choquet, Y., and Wollman, F.A. (2011). Dual functions of the nucleus-encoded factor TDA1 in trapping and translation activation of *atpA* transcripts in *Chlamydomonas reinhardtii* chloroplasts. Plant J **67,** 1055-1066.

Ebersold, W.T. (1967). *Chlamydomonas reinhardtii*: heterozygous diploid strains. Science **157,** 447-449.

Edgar, R.C. (2004). MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res **32,** 1792-1797.

Elkonin, L.A. (2005). Dominant male sterility in sorghum: effect of nuclear background on inheritance of tissue-culture-induced mutation. Theor Appl Genet **111,** 1377-1384.

Feng, S., Chen, Y., Kamada, K., Wang, H., Tang, K., Wang, M., and Gao, Y.G. (2013). YoeB-ribosome structure: a canonical RNase that requires the ribosome for its specific activity. Nucleic Acids Res **41,** 9549-9556.

Filipovska, A., and Rackham, O. (2011). Designer RNA-binding proteins: New tools for manipulating the transcriptome. RNA biology **8,** 978-983.

Filipovska, A., and Rackham, O. (2012). Modular recognition of nucleic acids by PUF, TALE and PPR proteins. Molecular bioSystems **8,** 699-708.

Filipovska, A., Razif, M.F., Nygard, K.K., and Rackham, O. (2011). A universal code for RNA recognition by PUF proteins. Nature chemical biology **7,** 425-427.

Fischer, N., Stampacchia, O., Redding, K., and Rochaix, J.D. (1996). Selectable marker recycling in the chloroplast. Mol Gen Genet **251,** 373-380.

Forner, J., Holzle, A., Jonietz, C., Thuss, S., Schwarzlander, M., Weber, B., Meyer, R.C., and Binder, S. (2008). Mitochondrial mRNA polymorphisms in different *Arabidopsis* accessions. Plant Physiol **148,** 1106-1116.

Fujii, S., Bond, C.S., and Small, I.D. (2011). Selection patterns on restorer-like genes reveal a conflict between nuclear and mitochondrial genomes throughout angiosperm evolution. Proc Natl Acad Sci U S A **108,** 1723-1728.

Geddy, R., and Brown, G.G. (2007). Genes encoding pentatricopeptide repeat (PPR) proteins are not conserved in location in plant genomes and may be subject to diversifying selection. BMC Genomics **8,** 130.

Germain, A., Hotto, A.M., Barkan, A., and Stern, D.B. (2013). RNA processing and decay in plastids. Wiley interdisciplinary reviews. RNA **4,** 295-316.

Goldschmidt-Clermont, M. (1991). Transgenic expression of aminoglycoside adenine transferase in the chloroplast: a selectable marker of site-directed transformation of *Chlamydomonas*. Nucleic Acids Res. **19,** 4083-4089.

Gross, C.H., Ranum, L.P., and Lefebvre, P.A. (1988). Extensive restriction fragment length polymorphisms in a new isolate of *Chlamydomonas reinhardtii*. Curr Genet **13,** 503-508.

Guindon, S., Dufayard, J.F., Lefort, V., Anisimova, M., Hordijk, W., and Gascuel, O. (2010). New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. Systematic biology **59,** 307-321.

Hammani, K., Bonnard, G., Bouchoucha, A., Gobert, A., Pinker, F., Salinas, T., and Giege, P. (2014). Helical repeats modular proteins are major players for organelle gene expression. Biochimie **100,** 141-150.

Harris, E.H. (1989). The *Chlamydomonas* source book: a comprehensive guide to biology and laboratory use. (San Diego: Academic Press).

Hernandez Mora, J.R., Rivals, E., Mireau, H., and Budar, F. (2010). Sequence analysis of two alleles reveals that intra-and intergenic recombination played a role in the evolution of the radish fertility restorer (Rfo). BMC Plant Biol **10,** 35. doi: 10.1186/1471-2229-10-35.

Higgs, D.C., Shapiro, R.S., Kindle, K.L., and Stern, D.B. (1999). Small cis-acting sequences that specify secondary structures in a chloroplast mRNA are essential for RNA stability and translation. Mol Cell Biol **19,** 8479-8491.

Higuchi, R. (1990). Recombinant PCR. In: PCR protocols: a guide to methods and application. (London, New York: Academic press).

Holzle, A., Jonietz, C., Torjek, O., Altmann, T., Binder, S., and Forner, J. (2011). A RESTORER OF FERTILITY-like PPR gene is required for 5'-end processing of the *nad4* mRNA in mitochondria of *Arabidopsis thaliana*. Plant J **65,** 737-744.

Jonietz, C., Forner, J., Holzle, A., Thuss, S., and Binder, S. (2010). RNA PROCESSING FACTOR2 is required for 5' end processing of *nad9* and *cox3* mRNAs in mitochondria of *Arabidopsis thaliana*. Plant Cell **22,** 443-453.

Kathir, P., LaVoie, M., Brazelton, W.J., Haas, N.A., Lefebvre, P.A., and Silflow, C.D. (2003). Molecular map of the *Chlamydomonas reinhardtii* nuclear genome. Eukaryot Cell **2,** 362-379.

Ke, J., Chen, R.Z., Ban, T., Zhou, X.E., Gu, X., Tan, M.H., Chen, C., Kang, Y., Brunzelle, J.S., Zhu, J.K., Melcher, K., and Xu, H.E. (2013). Structural basis for RNA recognition by a dimeric PPR-protein complex. Nat Struct Mol Biol **20,** 1377-1382.

Ketchner, S.L., Drapier, D., Olive, J., Gaudriault, S., Girard-Bascou, J., and Wollman, F.A. (1995). Chloroplasts can accommodate inclusion bodies. Evidence from a mutant of *Chlamydomonas reinhardtii* defective in the assembly of the chloroplast ATP synthase. J Biol Chem **270,** 15299-15306.

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.

Kuras, R., and Wollman, F.A. (1994). The assembly of cytochrome b_6f complexes: an approach using genetic transformation of the green alga *Chlamydomonas reinhardtii*. Embo J **13,** 1019-1027.

Lee, I., and Hong, W. (2004). RAP--a putative RNA-binding domain. Trends Biochem Sci **29,** 567-570.

Li, H., and Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics (Oxford, England) **25,** 1754-1760.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., and Durbin, R. (2009). The Sequence Alignment/Map format and SAMtools. Bioinformatics (Oxford, England) **25,** 2078-2079.

Loiselay, C. (2007). Importance of nucleus-encoded factorsfor the expression of the chloroplast genome: the example of the chhloroplast *petA* gene in *Chlamydomonas reinhardtii*/ Importance des facteurs codés par le génome nucléaire pour l'expression du génome chloroplastique: L'exemple du gène chloroplastique *petA* chez *Chlamydomonas reinhardtii*. PhD thesis, 218 p., UFR de Biologie, Université Pierre et Marie Curie, Paris;

Loiselay, C., Gumpel, N.J., Girard-Bascou, J., Watson, A.T., Purton, S., Wollman, F.A., and Choquet, Y. (2008). Molecular identification and function of cis- and trans-acting determinants for *petA* transcript stability in *Chlamydomonas reinhardtii* chloroplasts. Mol Cell Biol **28,** 5529-5542.

Lurin, C., Andres, C., Aubourg, S., Bellaoui, M., Bitton, F., Bruyere, C., Caboche, M., Debast, C., Gualberto, J., Hoffmann, B., Lecharny, A., Le Ret, M., Martin-Magniette, M.L., Mireau, H., Peeters, N., Renou, J.P., Szurek, B., Taconnat, L., and Small, I. (2004). Genome-wide analysis of *Arabidopsis* pentatricopeptide repeat proteins reveals their essential role in organelle biogenesis. Plant Cell **16,** 2089-2103.

Mak, A.N., Bradley, P., Cernadas, R.A., Bogdanove, A.J., and Stoddard, B.L. (2012). The crystal structure of TAL effector PthXo1 bound to its DNA target. Science **335,** 716-719.

Maxwell, K., and Johnson, G.N. (2000). Chlorophyll fluorescence--a practical guide. J Exp Bot **51,** 659-668.

Merchant, S.S., Prochnik, S.E., Vallon, O., Harris, E.H., Karpowicz, S.J., Witman, G.B., Terry, A., Salamov, A., Fritz-Laylin, L.K., Marechal-Drouard, L., Marshall, W.F., Qu, L.H., Nelson, D.R., Sanderfoot, A.A., Spalding, M.H., Kapitonov, V.V., Ren, Q., Ferris, P., Lindquist, E., Shapiro, H., Lucas, S.M., Grimwood, J., Schmutz, J., Cardol, P., Cerutti, H., Chanfreau, G., Chen, C.L., Cognat, V., Croft, M.T., Dent, R., Dutcher, S., Fernandez, E., Fukuzawa, H., Gonzalez-Ballester, D., Gonzalez-Halphen, D., Hallmann, A., Hanikenne, M., Hippler, M., Inwood, W., Jabbari, K., Kalanon, M., Kuras, R., Lefebvre, P.A., Lemaire, S.D., Lobanov, A.V., Lohr, M., Manuell, A., Meier, I., Mets, L., Mittag, M., Mittelmeier, T., Moroney, J.V., Moseley, J., Napoli, C., Nedelcu, A.M., Niyogi, K., Novoselov, S.V., Paulsen, I.T., Pazour, G., Purton, S., Ral, J.P., Riano-Pachon, D.M., Riekhof, W., Rymarquis, L., Schroda, M., Stern, D., Umen, J.,

Willows, R., Wilson, N., Zimmer, S.L., Allmer, J., Balk, J., Bisova, K., Chen, C.J., Elias, M., Gendler, K., Hauser, C., Lamb, M.R., Ledford, H., Long, J.C., Minagawa, J., Page, M.D., Pan, J., Pootakham, W., Roje, S., Rose, A., Stahlberg, E., Terauchi, A.M., Yang, P., Ball, S., Bowler, C., Dieckmann, C.L., Gladyshev, V.N., Green, P., Jorgensen, R., Mayfield, S., Mueller-Roeber, B., Rajamani, S., Sayre, R.T., Brokstein, P., Dubchak, I., Goodstein, D., Hornick, L., Huang, Y.W., Jhaveri, J., Luo, Y., Martinez, D., Ngau, W.C., Otillar, B., Poliakov, A., Porter, A., Szajkowski, L., Werner, G., Zhou, K., Grigoriev, I.V., Rokhsar, D.S., and Grossman, A.R. (2007). The *Chlamydomonas* genome reveals the evolution of key animal and plant functions. Science **318,** 245-250.

Michelmore, R.W., and Meyers, B.C. (1998). Clusters of resistance genes in plants evolve by divergent selection and a birth-and-death process. Genome Res **8,** 1113-1130.

Miller, M.T., Higgin, J.J., and Hall, T.M. (2008). Basis of altered RNA-binding specificity by PUF proteins revealed by crystal structures of yeast Puf4p. Nat Struct Mol Biol **15,** 397-402.

Moffett, P. (2009). Mechanisms of recognition in dominant R gene mediated resistance. Adv Virus Res **75,** 1-33.

Monod, C., Goldschmidt-Clermont, M., and Rochaix, J.D. (1992). Accumulation of chloroplast *psbB* RNA requires a nuclear factor in *Chlamydomonas reinhardtii*. Mol Gen Genet **231,** 449-459.

Moscou, M.J., and Bogdanove, A.J. (2009). A simple cipher governs DNA recognition by TAL effectors. Science **326,** 1501.

Neubauer, C., Gao, Y.G., Andersen, K.R., Dunham, C.M., Kelley, A.C., Hentschel, J., Gerdes, K., Ramakrishnan, V., and Brodersen, D.E. (2009). The structural basis for mRNA recognition and cleavage by the ribosome-dependent endonuclease RelE. Cell **139,** 1084-1095.

Odaert, B., Saida, F., Aliprandi, P., Durand, S., Crechet, J.B., Guerois, R., Laalami, S., Uzan, M., and Bontems, F. (2007). Structural and functional studies of RegB, a new member of a family of sequence-specific ribonucleases involved in mRNA inactivation on the ribosome. J Biol Chem **282,** 2019-2028.

Qi, D., and Innes, R.W. (2013). Recent Advances in Plant NLR Structure, Function, Localization, and Signaling. Frontiers in immunology **4,** 348.

Qi, J., and Zhao, F. (2011). inGAP-sv: a novel scheme to identify and visualize structural variation from paired end mapping data. Nucleic Acids Res **39,** W567-575.

Qi, J., Zhao, F., Buboltz, A., and Schuster, S.C. (2010). inGAP: an integrated next-generation genome analysis pipeline. Bioinformatics (Oxford, England) **26,** 127-129.

Quinlan, A.R., and Hall, I.M. (2010). BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics (Oxford, England) **26,** 841-842.

Rahire, M., Laroche, F., Cerutti, L., and Rochaix, J.D. (2012). Identification of an OPR protein involved in the translation initiation of the PsaB subunit of photosystem I. Plant J **72,** 652-661.

Raynaud, C., Loiselay, C., Wostrikoff, K., Kuras, R., Girard-Bascou, J., Wollman, F.A., and Choquet, Y. (2007). Evidence for regulatory function of nucleus-encoded factors on mRNA stabilization and translation in the chloroplast. Proc Natl Acad Sci U S A **104,** 9093-9098.

Robinson, J.T., Thorvaldsdottir, H., Winckler, W., Guttman, M., Lander, E.S., Getz, G., and Mesirov, J.P. (2011). Integrative genomics viewer. Nat Biotechnol **29,** 24-26.

Rubinson, E.H., and Eichman, B.F. (2012). Nucleic acid recognition by tandem helical repeats. Curr Opin Struct Biol **22,** 101-109.

Rymarquis, L.A., Handley, J.M., Thomas, M., and Stern, D.B. (2005). Beyond complementation. Map-based cloning in *Chlamydomonas reinhardtii*. Plant Physiol **137,** 557-566.

Sakamoto, W., Chen, X., Kindle, K.L., and Stern, D.B. (1994). Function of the *Chlamydomonas reinhardtii petD* 5' untranslated region in regulating the accumulation of subunit IV of the cytochrome b_6f complex. Plant J. **6,** 503-512.

Sambrook, J., Fritsch, E.F., and Maniatis, T. (1989). Molecular Cloning. (Cold Spring Harbor Laboratory Press).

Schmitz-Linneweber, C., and Small, I. (2008). Pentatricopeptide repeat proteins: a socket set for organelle gene expression. Trends Plant Sci **13,** 663-670.

Schneeberger, K., Ossowski, S., Lanz, C., Juul, T., Petersen, A.H., Nielsen, K.L., Jorgensen, J.E., Weigel, D., and Andersen, S.U. (2009). SHOREmap: simultaneous mapping and mutation identification by deep sequencing. Nature methods **6,** 550-551.

Seeholzer, S., Tsuchimatsu, T., Jordan, T., Bieri, S., Pajonk, S., Yang, W., Jahoor, A., Shimizu, K.K., Keller, B., and Schulze-Lefert, P. (2010). Diversity at the Mla powdery mildew resistance locus from cultivated barley reveals sites of positive selection. Mol Plant Microbe Interact **23,** 497-509.

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.

Simpson, J.T., McIntyre, R.E., Adams, D.J., and Durbin, R. (2010). Copy number variant detection in inbred strains from short read sequence data. Bioinformatics (Oxford, England) **26,** 565-567.

Sizova, I., Fuhrmann, M., and Hegemann, P. (2001). A *Streptomyces rimosus aphVIII* gene coding for a new type phosphotransferase provides stable antibiotic resistance to *Chlamydomonas reinhardtii*. Gene **277,** 221-229.

Sizova, I.A., Lapina, T.V., Frolova, O.N., Alexandrova, N.N., Akopiants, K.E., and Danilenko, V.N. (1996). Stable nuclear transformation of *Chlamydomonas reinhardtii* with a *Streptomyces rimosu*s gene as the selective marker. Gene **181,** 13-18.

Slater, G.S., and Birney, E. (2005). Automated generation of heuristics for biological sequence comparison. BMC Bioinformatics **6,** 31.

Song, L.Q., Fu, T.D., Tu, J.X., Ma, C.Z., and Yang, G.S. (2006). Molecular validation of multiple allele inheritance for dominant genic male sterility gene in *Brassica napus L*. Theor Appl Genet **113,** 55-62.

Suyama, M., Torrents, D., and Bork, P. (2006). PAL2NAL: robust conversion of protein sequence alignments into the corresponding codon alignments. Nucleic Acids Res **34,** W609-612.

Tamura, K., Stecher, G., Peterson, D., Filipski, A., and Kumar, S. (2013). MEGA6: Molecular Evolutionary Genetics Analysis version 6.0. Mol Biol Evol **30,** 2725-2729.

Thorvaldsdottir, H., Robinson, J.T., and Mesirov, J.P. (2013). Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. Briefings in bioinformatics **14,** 178-192.

Tourasse, N.J., Choquet, Y., and Vallon, O. (2013). PPR proteins of green algae. RNA biology **10,** 1526-1542.

Vaistij, F.E., Goldschmidt-Clermont, M., Wostrikoff, K., and Rochaix, J.D. (2000). Stability determinants in the chloroplast *psbB/T/H* mRNAs of *Chlamydomonas reinhardtii*. Plant J **21,** 469-482.

Wang, X., Zamore, P.D., and Hall, T.M. (2001). Crystal structure of a Pumilio homology domain. Mol Cell **7,** 855-865.

Wang, X., McLachlan, J., Zamore, P.D., and Hall, T.M. (2002). Modular recognition of RNA by a human pumilio-homology domain. Cell **110,** 501-512.

Werner, R., and Mergenhagen, D. (1998). Mating Type Determination of *Chlamydomonas reinhardtii* by PCR. Plant Mol Biol Report **16,** 295-299.

Wong, K., Keane, T.M., Stalker, J., and Adams, D.J. (2010). Enhanced structural variant and breakpoint detection using SVMerge by integration of multiple detection methods and local assembly. Genome Biol **11,** R128.

Woodson, J.D., and Chory, J. (2008). Coordination of gene expression between organellar and nuclear genomes. Nat Rev Genet **9,** 383-395.

Wostrikoff, K., Choquet, Y., Wollman, F.A., and Girard-Bascou, J. (2001). TCA1, a single nuclear-encoded translational activator specific for *petA* mRNA in *Chlamydomonas reinhardtii* chloroplast. Genetics **159,** 119-132.

Wostrikoff, K., Girard-Bascou, J., Wollman, F.A., and Choquet, Y. (2004). Biogenesis of PSI involves a cascade of translational autoregulation in the chloroplast of *Chlamydomonas*. Embo J **23,** 2696-2705.

Wulff, B.B., Heese, A., Tomlinson-Buhot, L., Jones, D.A., de la Pena, M., and Jones, J.D. (2009). The major specificity-determining amino acids of the tomato Cf-9 disease resistance protein are at hypervariable solvent-exposed positions in the central leucine-rich repeats. Mol Plant Microbe Interact **22,** 1203-1213.

Xu, B., and Yang, Z. (2013). PAMLX: a graphical user interface for PAML. Mol Biol Evol **30,** 2723-2724.

Yagi, Y., Nakamura, T., and Small, I. (2014). The potential for manipulating RNA with pentatricopeptide repeat proteins. Plant J **78,** 772-782.

Yagi, Y., Hayashi, S., Kobayashi, K., Hirayama, T., and Nakamura, T. (2013). Elucidation of the RNA recognition code for pentatricopeptide repeat proteins involved in organelle RNA editing in plants. PLoS One **8,** e57286.

Yang, Z. (2007). PAML 4: phylogenetic analysis by maximum likelihood. Mol Biol Evol **24,** 1586-1591.

Yang, Z., and Swanson, W.J. (2002). Codon-substitution models to detect adaptive evolution that account for heterogeneous selective pressures among site classes. Mol Biol Evol **19,** 49-57.

Ye, K., Schulz, M.H., Long, Q., Apweiler, R., and Ning, Z. (2009). Pindel: a pattern growth approach to detect break points of large deletions and medium sized insertions from paired-end short reads. Bioinformatics (Oxford, England) **25,** 2865-2871.

Ye, Z., and Ting, J.P. (2008). NLR, the nucleotide-binding domain leucine-rich repeat containing gene family. Current opinion in immunology **20,** 3-9.

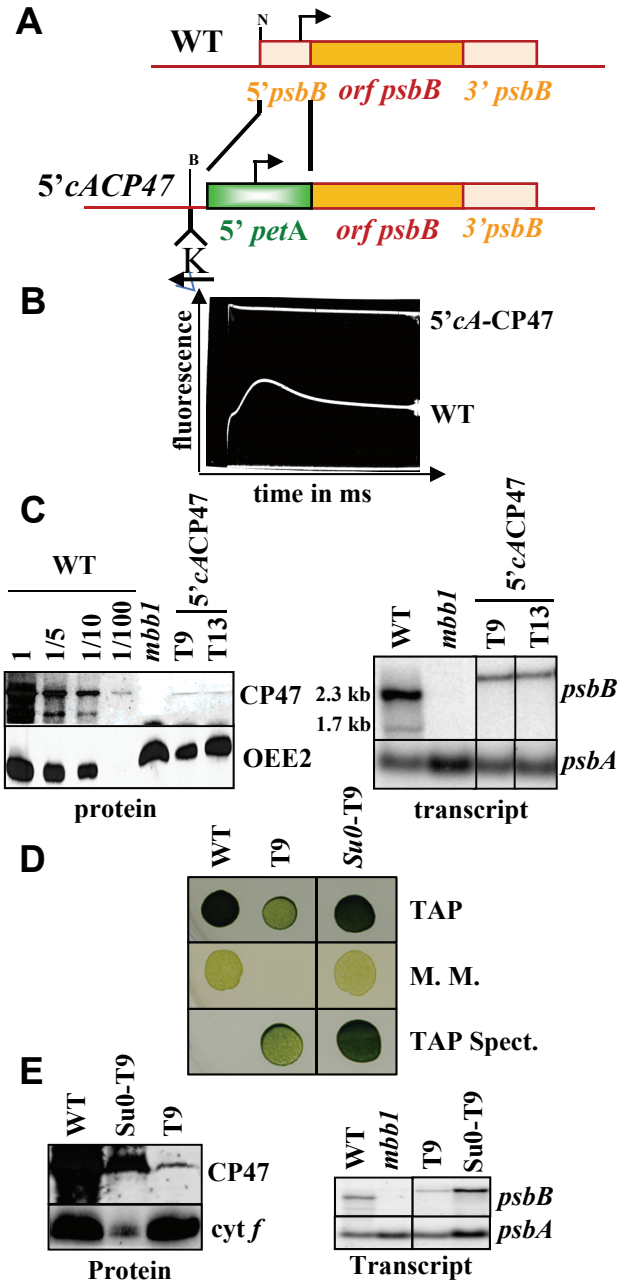
Yin, P., Li, Q., Yan, C., Liu, Y., Liu, J., Yu, F., Wang, Z., Long, J., He, J., Wang, H.W., Wang, J., Zhu, J.K., Shi, Y., and Yan, N. (2013). Structural basis for the modular recognition of single-stranded RNA by PPR proteins. Nature **504,** 168-171.

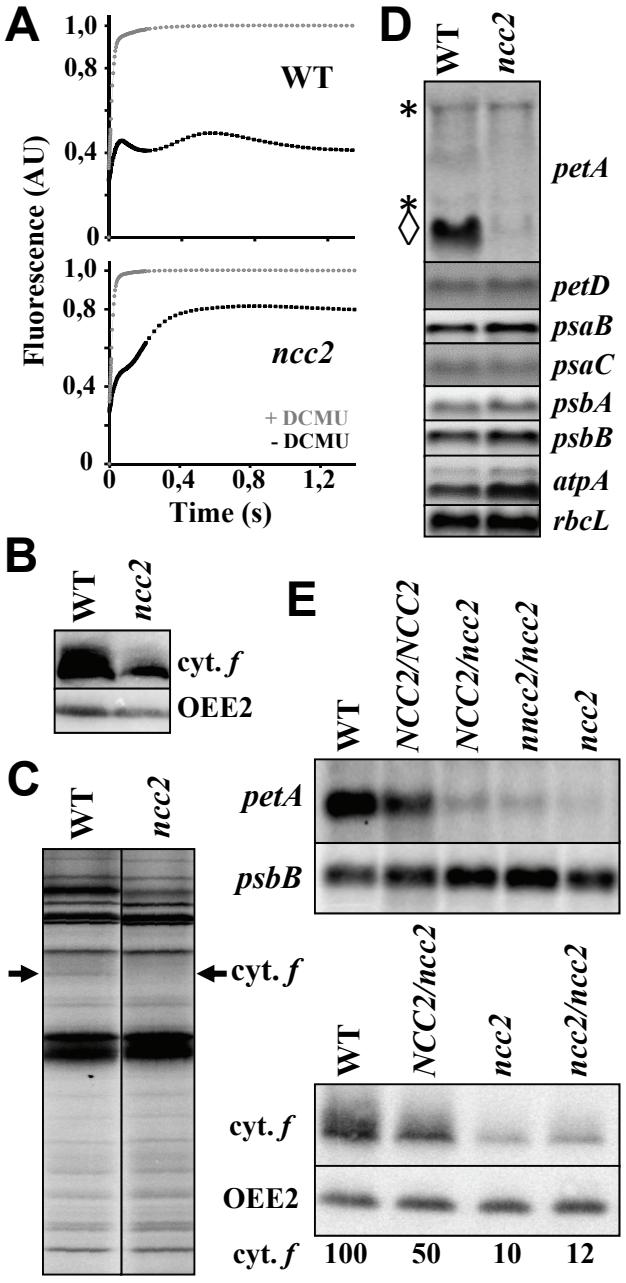
Zamora, I., Feldman, J.L., and Marshall, W.F. (2004). PCR-based assay for mating type and diploidy in *Chlamydomonas*. Biotechniques **37,** 534-536.

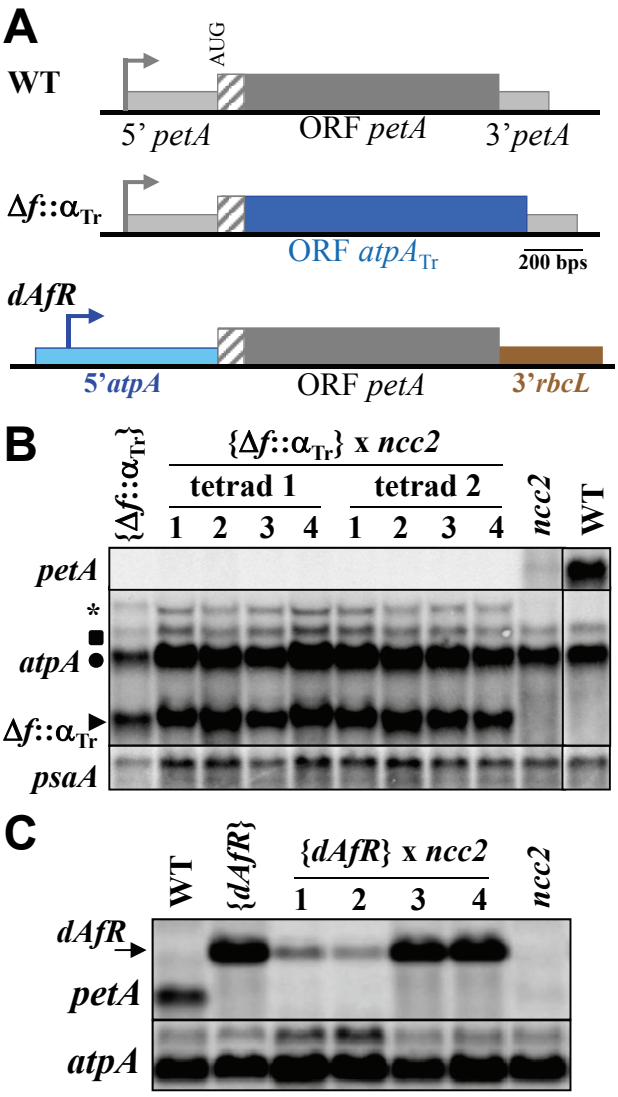
Zerbino, D.R., and Birney, E. (2008). Velvet: algorithms for de novo short read assembly using de Bruijn graphs. Genome Res **18,** 821-829.

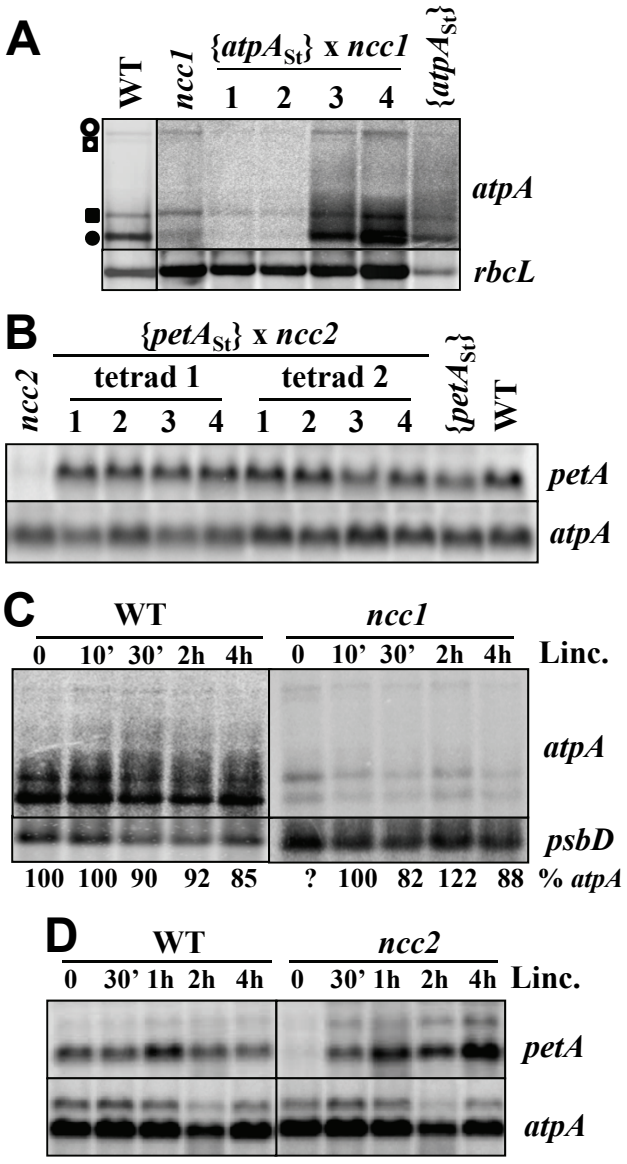
Zhang, Y. (2008). I-TASSER server for protein 3D structure prediction. BMC Bioinformatics **9,** 40.

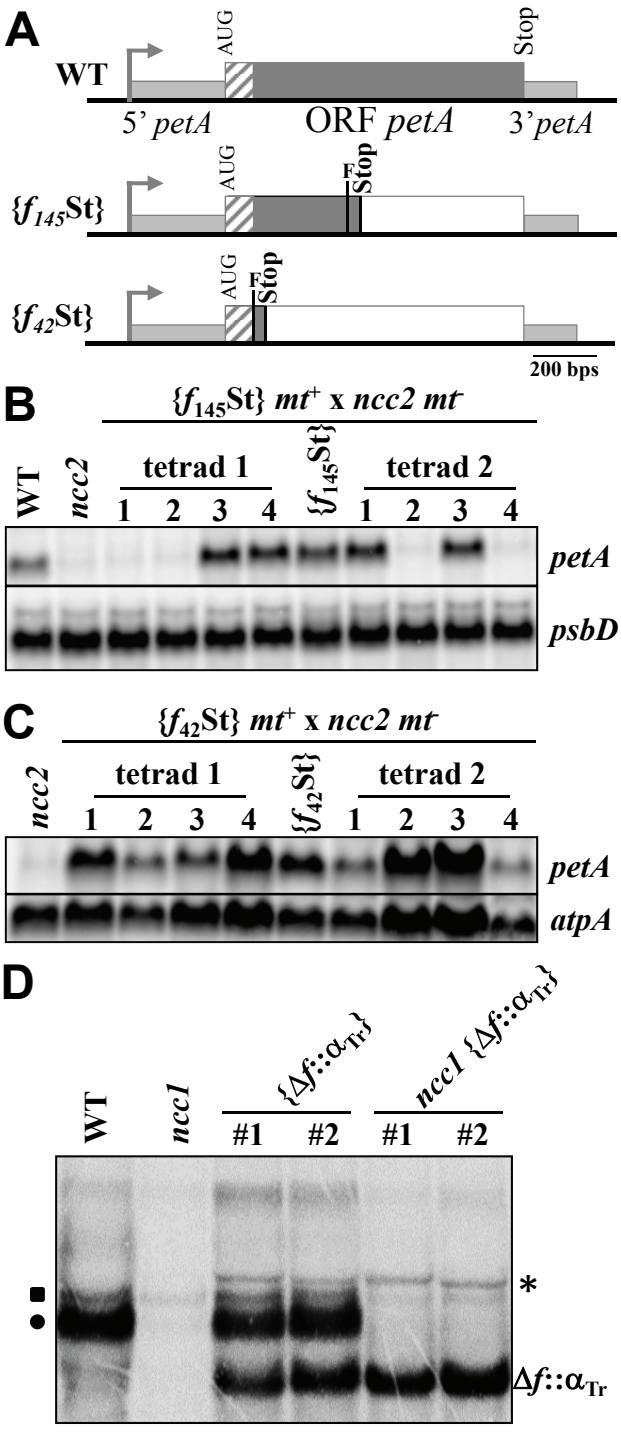
Zito, F., Kuras, R., Choquet, Y., Kossel, H., and Wollman, F.A. (1997). Mutations of cytochrome b_6 in *Chlamydomonas reinhardtii* disclose the functional significance for a proline to leucine conversion by *petB* editing in maize and tobacco. Plant Mol Biol **33,** 79-86.

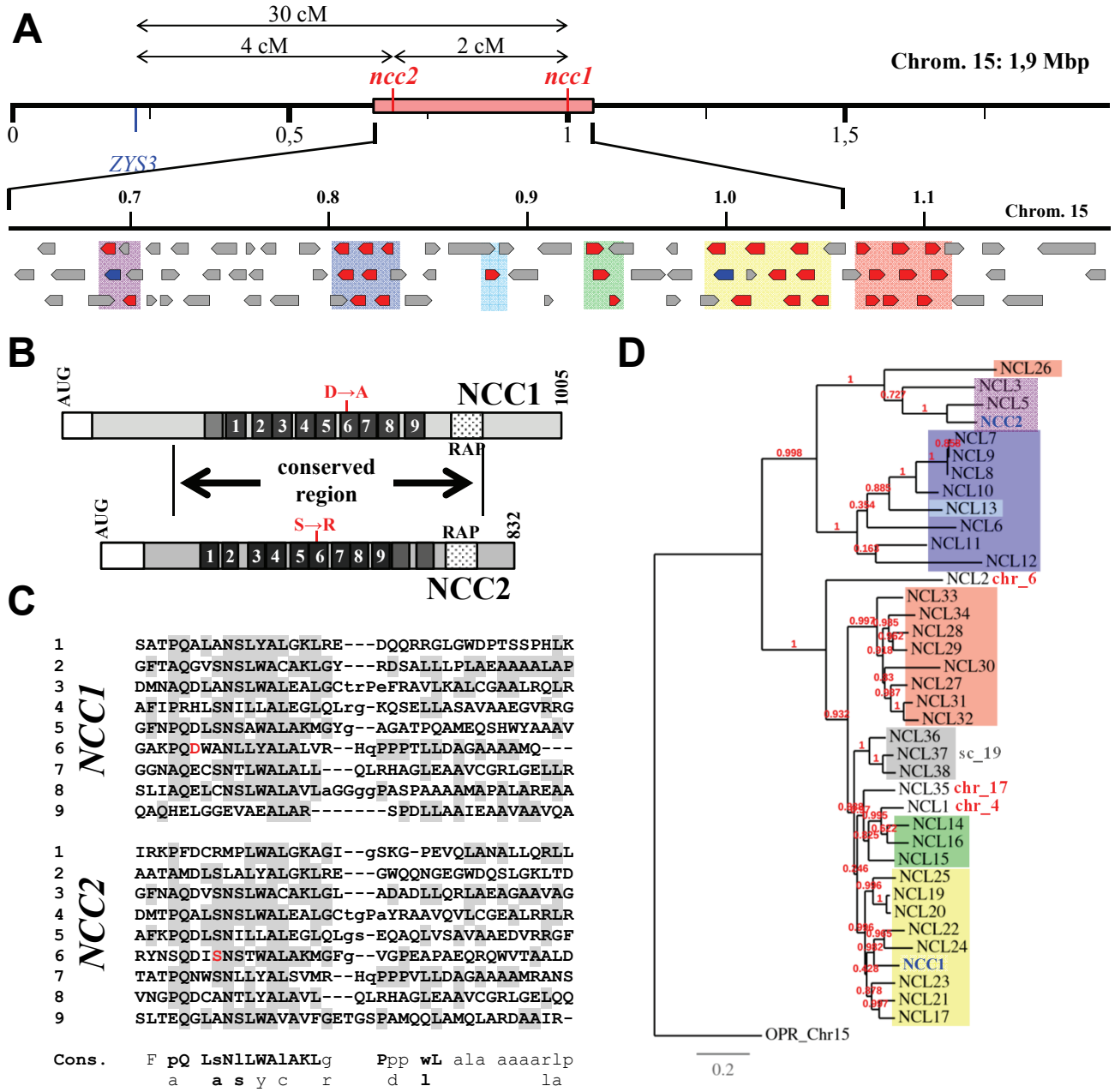

A
WT
N
5'psbB orf psbB 3' psbB
5'cACP47
B
5' petA orf psbB 3'psbB
K
B
fluorescence
5'cA-CP47
WT
time in ms
C
WT
5'cACP47
1
1/5
1/10
1/100
mbb1
T9
T13
CP47
OEE2
protein
WT
mbb1
5'cACP47
T9
T13
2.3 kb
1.7 kb
psbB
psbA
transcript
D
WT
T9
Su0-T9
TAP
M. M.
TAP Spect.
E
WT
Su0-T9
T9
CP47
cyt f
Protein
WT
mbb1
T9
Su0-T9
psbB
psbA
Transcript

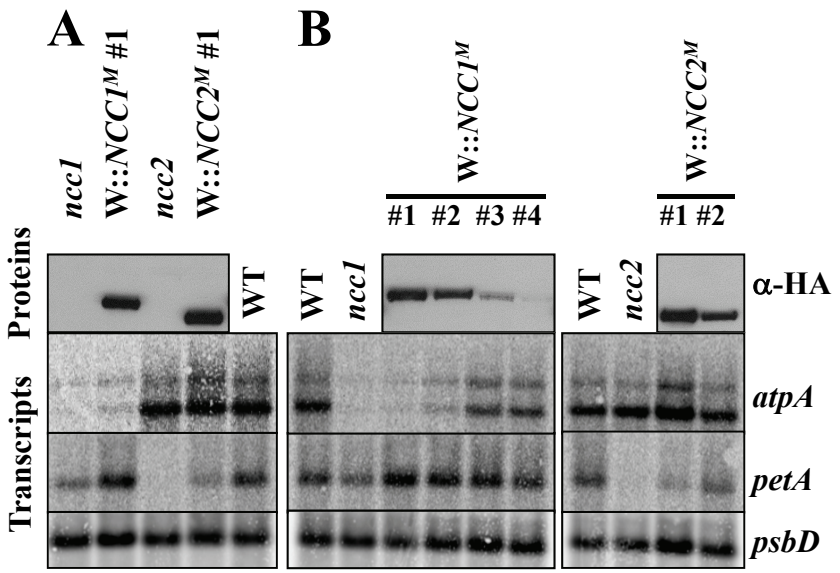

A
1,0
0,8
0,4
0
WT
Fluorescence (AU)
ncc2
+ DCMU
- DCMU
0
0,4
0,8
1,2
Time (s)
B
WT
ncc2
cyt. f
OEE2
C
WT
ncc2
cyt. f
D
WT
ncc2
petA
petD
psaB
psaC
psbA
psbB
atpA
rbcL
E
WT
NCC2/NCC2
NCC2/ncc2
nncc2/ncc2
ncc2
petA
psbB
WT
NCC2/ncc2
ncc2
ncc2/ncc2
cyt. f
OEE2
cyt. f
100
50
10
12

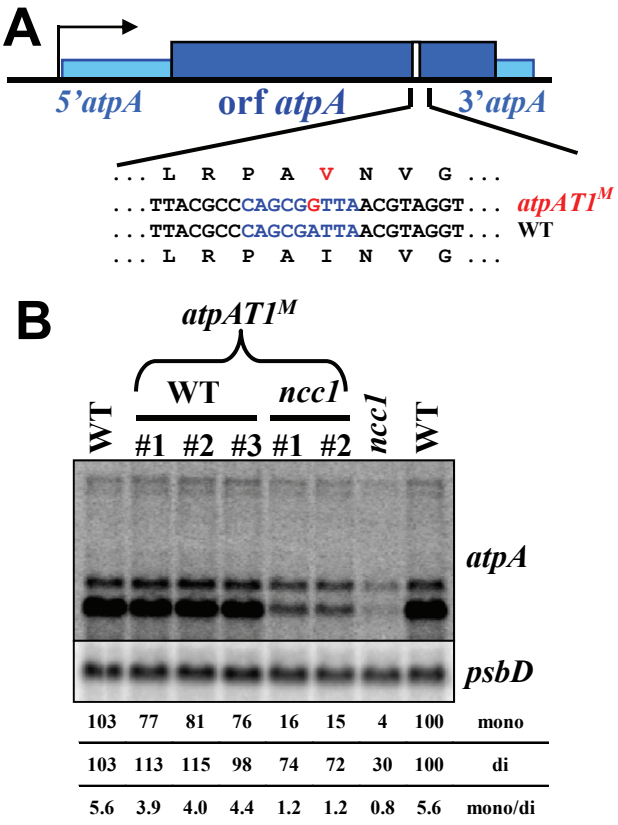

A
WT
AUG
5' petA
ORF petA
3'petA
Δf::αTr
ORF atpATr
200 bps
dAfR
5'atpA
ORF petA
3'rbcL
B
{Δf::αTr}
{Δf::αTr} x ncc2
tetrad 1
tetrad 2
1 2 3 4 1 2 3 4
ncc2
WT
petA
atpA
Δf::αTr
psaA
C
WT
{dAfR}
{dAfR} x ncc2
1 2 3 4
ncc2
dAfR
petA
atpA

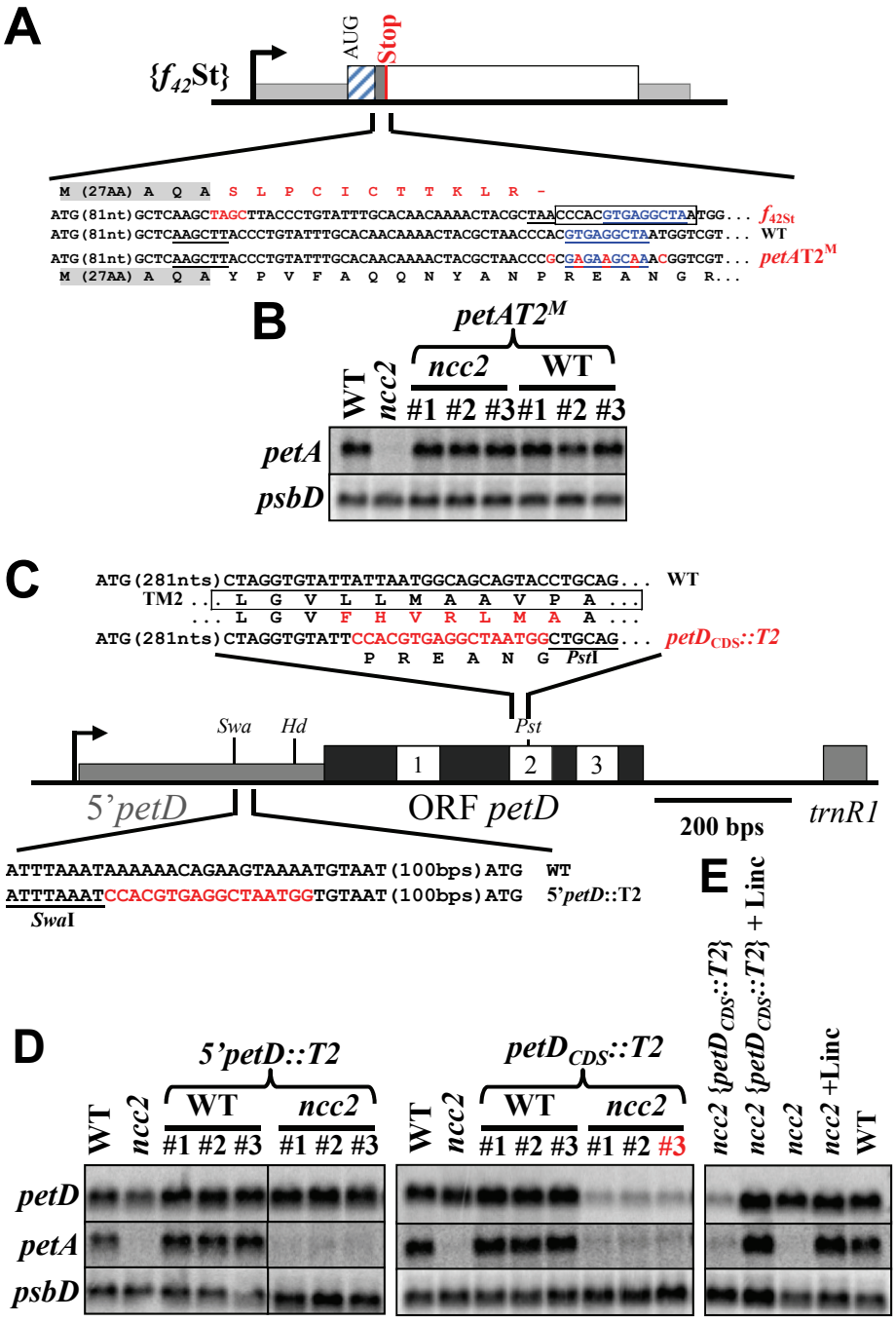

A
WT
ncc1
{atpA_St} x ncc1
1 2 3 4
{atpA_St}
atpA
rbcL
B
{petA_St} x ncc2
tetrad 1
tetrad 2
ncc2
1 2 3 4 1 2 3 4
{petA_St}
WT
petA
atpA
C
WT
ncc1
0 10' 30' 2h 4h 0 10' 30' 2h 4h
Linc.
atpA
psbD
100 100 90 92 85 ? 100 82 122 88 % atpA
D
WT
ncc2
0 30' 1h 2h 4h 0 30' 1h 2h 4h
Linc.
petA
atpA


A
WT
AUG
Stop
5' petA
ORF petA
3'petA
{f145St}
F
{f42St}
200 bps
B
{f145St} mt+ x ncc2 mt-
WT
ncc2
tetrad 1
tetrad 2
1 2 3 4
petA
psbD
C
{f42St} mt+ x ncc2 mt-
petA
atpA
D
ncc1
{Δf::αTr}
#1 #2
ncc1 {Δf::αTr}
*
Δf::αTr


A
30 cM
4 cM
2 cM
ncc2
ncc1
Chrom. 15: 1,9 Mbp
0
0,5
1
1,5
ZYS3
0.7
0.8
0.9
1.0
1.1
Chrom. 15
B
AUG
D→A
NCC1
1005
RAP
conserved region
AUG
S→R
RAP
832
NCC2
C
NCC1
1 SATPQALANSLYALGKLRE---DQQRRGLGWDPTSSPHLK
2 GFTAQGVSNSLWACAKLGY---RDSALLLPLAEAAAALAP
3 DMNAQDLANSLWALEALGCtrPeFRAVLKALCGAALRQLR
4 AFIPRHLSNILLALEGLQLrg-KQSELLASAVAAEGVRRG
5 GFNPQDLSNSAWALAKMGYg--AGATPQAMEQSHWYAAAV
6 GAKPQDWANLLYALALVR--HqPPPTLLDAGAAAAMQ---
7 GGNAQECSNTLWALALL---QLRHAGLEAAVCGRLGELLR
8 SLIAQELCNSLWALAVLaGGggPASPAAAAMAPALAREAA
9 QAQHELGGEVAEALAR-------SPDLLAAIEAAVAAVQA
NCC2
1 IRKPFDCRMPLWALGKAGI--gSKG-PEVQLANALLQRLL
2 AATAMDLSLALYALGKLRE---GWQQNGEGWDQSLGKLTD
3 GFNAQDVSNSLWACAKLGL---ADADLLQRLAEAGAAVAG
4 DMTPQALSNSLWALEALGCtgPaYRAAVQVLCGEALRRLR
5 AFKPQDLSNILLALEGLQLgs-EQAQLVSAVAAEDVRRGF
6 RYNSQDISNSTWALAKMGFg--VGPEAPAEQRQWVTAALD
7 TATPQNWSNLLYALSVMR--HqPPPVLLDAGAAAAMRANS
8 VNGPQDCANTLYALAVL---QLRHAGLEAAVCGRLGELQQ
9 SLTEQGLANSLWAVAVFGETGSPAMQQLAMQLARDAAIR-
Cons. F pQ LsNlLWAlAKLg Ppp wL ala aaaarlp
a a s y c r d l la
D
NCL26
NCL3
NCL5
NCC2
NCL7
NCL9
NCL8
NCL10
NCL13
NCL6
NCL11
NCL12
NCL2 chr_6
NCL33
NCL34
NCL28
NCL29
NCL30
NCL27
NCL31
NCL32
NCL36
NCL37 sc_19
NCL38
NCL35 chr_17
NCL1 chr_4
NCL14
NCL16
NCL15
NCL25
NCL19
NCL20
NCL22
NCL24
NCC1
NCL23
NCL21
NCL17
OPR_Chr15
0.2


A
B
ncc1
W::NCC1M #1
ncc2
W::NCC2M #1
W::NCC1M
#1 #2 #3 #4
W::NCC2M
#1 #2
Proteins
Transcripts
WT
WT
ncc1
WT
ncc2
α-HA
atpA
petA
psbD


A
5'atpA
orf atpA
3'atpA
... L R P A V N V G ...
...TTACGCCCAGCGGTTAACGTAGGT... atpAT1M
...TTACGCCCAGCGATTAACGTAGGT... WT
... L R P A I N V G ...
B
atpAT1M
WT
ncc1
WT #1 #2 #3 #1 #2 ncc1 WT
atpA
psbD
103 77 81 76 16 15 4 100 mono
103 113 115 98 74 72 30 100 di
5.6 3.9 4.0 4.4 1.2 1.2 0.8 5.6 mono/di


A
{f42St}
AUG
Stop
M (27AA) A Q A S L P C I C T T K L R -
ATG (81nt) GCTCAAGCTAGCTTACCCTGTATTTGCACAACAAAACTACGCTAACCCACGTGAGGCTAATGG... f42St
ATG (81nt) GCTCAAGCTTACCCTGTATTTGCACAACAAAACTACGCTAACCCACGTGAGGCTAATGGTCGT... WT
ATG (81nt) GCTCAAGCTTACCCTGTATTTGCACAACAAAACTACGCTAACCCGCGAGAAGCAAACGGTCGT... petAT2M
M (27AA) A Q A Y P V F A Q Q N Y A N P R E A N G R...
B
petAT2M
ncc2
WT
WT
ncc2
#1 #2 #3 #1 #2 #3
petA
psbD
C
ATG (281nts) CTAGGTGTATTATTAATGGCAGCAGTACCTGCAG... WT
TM2 ..L G V L L M A A V P A...
...L G V F H V R L M A A...
ATG (281nts) CTAGGTGTATTCCACGTGAGGCTAATGGCTGCAG... petDCDS::T2
P R E A N G PstI
Swa
Hd
Pst
1
2
3
5'petD
ORF petD
200 bps
trnR1
ATTTAAATAAAAAACAGAAGTAAAATGTAAT (100bps) ATG WT
ATTTAAATCCACGTGAGGCTAATGGTGTAAT (100bps) ATG 5'petD::T2
SwaI
D
5'petD::T2
petDCDS::T2
WT
ncc2
WT
ncc2
WT
ncc2
WT
ncc2
#1 #2 #3 #1 #2 #3
#1 #2 #3 #1 #2 #3
petD
petA
psbD
E
ncc2 {petDCDS::T2}
ncc2 {petDCDS::T2} + Linc
ncc2
ncc2 +Linc
WT

Supplemental information avalaible online:

- Fig. S1 Phenotype of the double mutant *ncc1 ncc2* {FAFA}, descendant of the cross *ncc1 mt*$^{+}$ {FAFA} x *ncc2 mt*$^{-}$, whose genome was sequenced.

- Fig. S2: alignment of NCL proteins.

The RAP domain may show endonucleolytic activity

- Fig. S3: I-TASSER alignment used for threading of the NCC1 RAP domain.
- Fig. S4: Comparison of NCC1 model with known structures of endonucleases

- Table SI: Genetic independence of the ncc2 mutation from MCA1 and TCA1 genes

- Table SII: markers designed to map the *ncc2* mutation.

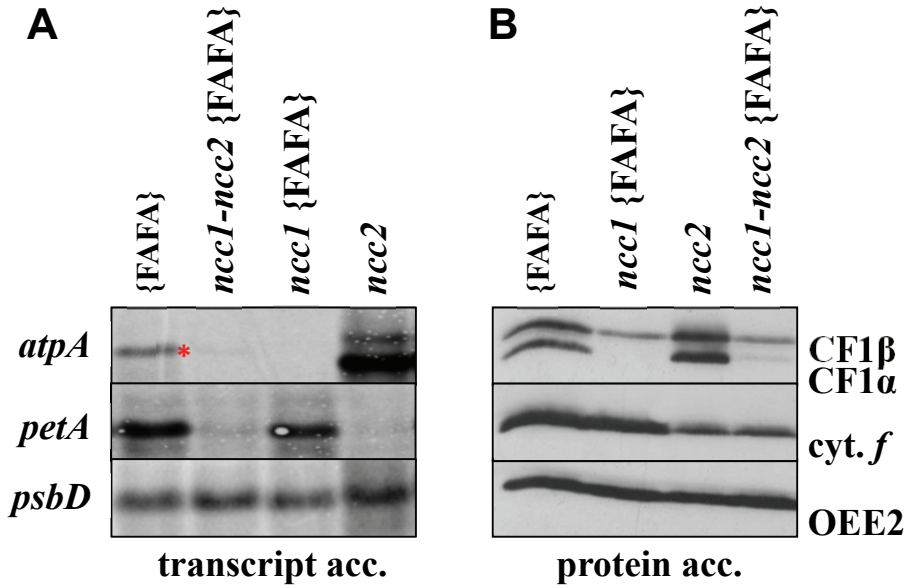
- Table SIII: oligonucleotides used in this study.

- Table SIV: BLAST search-based identification of NCL proteins in Phytozome v5.5.

- Table SV: description of NCL proteins.
  Improved gene models for *NCL7, 8, 21, 30, 35*
  Intracellular targeting of NCL proteins.

- Supplementary Material and Methods.
  Sequence of the synthetic *NCC1*^{M} and *NCC2*^{M} genes.

- Bibliography for Supplemental Information.

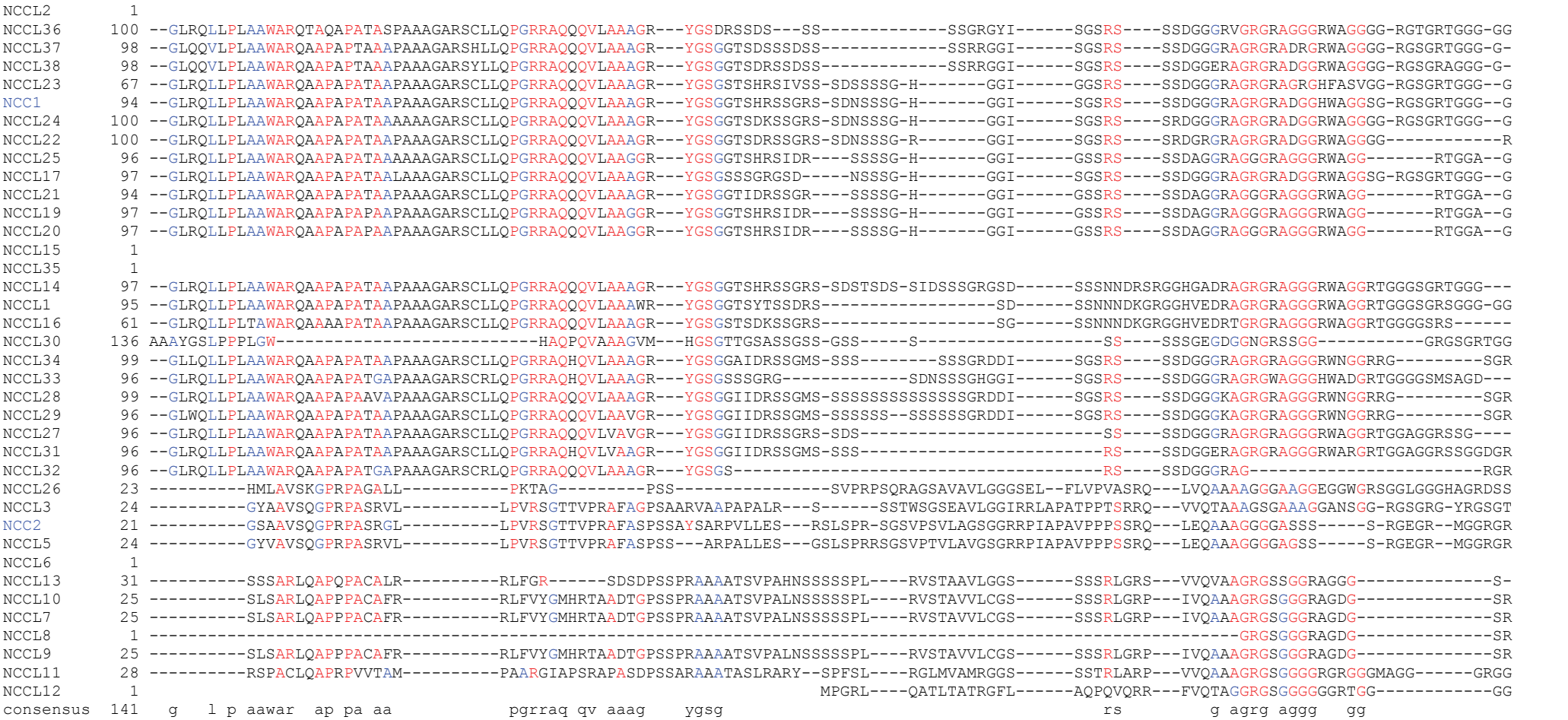


Supplemental Figure S1: **Phenotype of the double mutant *ncc1 ncc2* {FAFA}, descendant of the cross *ncc1 mt*+ {FAFA} x *ncc2 mt-*.**

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.

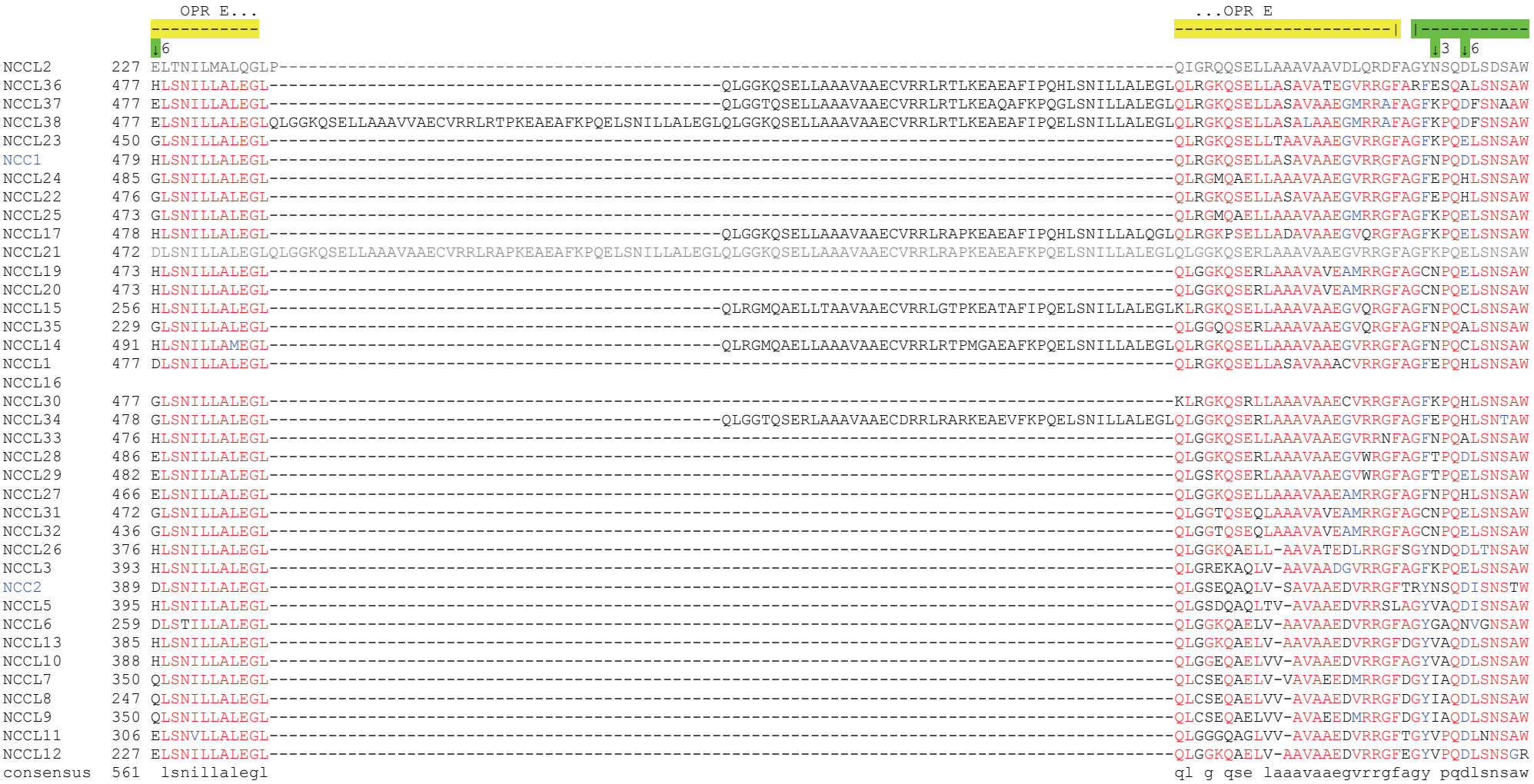
B. Accumulation of CF1-subunits α and β (detected using an antibody raised against the whole CF1 complex (Lemaire and Wollman, 1989)) and of cytochrome *f* (OEE2 serves as loading control) in the same strains.


NCCL2 1
NCCL36 1 MFQFAHG--------SAASGYGLGSRYNA----PASANTGL---NLCPCGIPQTQPRAARTAGRLGRLAPSAT----------TGGVARGLACSSATSYPAAAVLSGGGPHLAAAAPSATG AGG---
NCCL37 1 MLQFAHP--------SPANGHGLGSSYTS---APASPLIWH---RMCPCGIPQTQPRAVRTAKRL---GPSAT----------TGGVARGLACSSATSGPAAAVLSGGGPRLAAAAPAVTG AGG---
NCCL38 1 MLQFAHP--------SPANGHGLGSSYTS---APASPLIWH---RMCPCGIPQTQPRAVRTAKRL---GPSAT----------TGGVARGLACSSATSGPAAAVLSGGGPLLAAAAPAVTG AGG---
NCCL23 1 ----------MCPCGIPQTQPRAARTTGRL---GPSAT----------AGRVARGLACSSATSGPAAAVLSSSSPHLAAATPAATW AGG---
NCC1 1 MLQFAHGSA------VG-----LR-RSYN---APASAHIGL---ELWPCGIRLTQPRAARTAGRL---GPSAT----------AGRVARGLACSSATSGPAAAVLSGGGPRLAATAPAATG AGG---
NCCL24 1 MLQFTRASA------AAASGHDLGSRYTS---APASALVWH---RVCPCGIPLTQPRAARTAGRL---GPSAT----------AGGVARGLACSSATSGPAAAVLSSSSPHLAAATPAATW AGG---
NCCL22 1 MLQFTRASA------AAASGHDLGSRYTS---APASALVWH---RVCPCGIPLTQPRVARTTGRL---GPSAT----------AGGVARGLACSSATSGPAAAVLSSSSPHLAAATPAATW AGG---
NCCL25 1 ML-LFTRAS------AAASGHDLGCRYTS---APASALIWH---RVCPSGIPQS---AARTAGRL---GPSAT----------AGGLARGLACSSATSGPAAAVLSGGGPRLAAATPAATG AGG---
NCCL17 1 MLQ--SGRG------TAASRHGLGSSYN----ASASVRTWH---RVGPCGIPLTQPRAARTAGRL---GPSAT----------AGGVARGLACSSATSGPAAAVLSGGGPRLAAAAPAATG AGG---
NCCL21 1 MLQFAHGSAV-----------GLRRSYN----APASAHIGL---ELCPFGIRLTQPRAARTAGRL---GPSAT----------AGGVARGLACSSATSGPAAAVLSDGGPLLPATAPAATG AGG---
NCCL19 1 MLQ--SGRG------TAASRHGLGSSYN----APASARTWH---RVGPCGIPLTQPRAARTAGRL---GPSAT----------AGGVARGLACSSATSGPAAAVLSEGGPLLPATAPAATG AGG---
NCCL20 1 MLQ--SGRG------TAASRHGLGSSYN----APASARTWH---RVGPCGIPLTQPRAARTAGRL---GPSAT----------AGGVARGLACSSATSGPAAAVLSEGGPLLPATAPAATG AGG---
NCCL15 1
NCCL35 1
NCCL14 1 ML--RAGRG------SAASGHGLGSSYN----APASARTWH---RMCPCGTPLTQPRAARTAGRL---GPSAT----------AGGLARGLTCSSATSGPAAAVLSGGFPRLAATAPAATG AGG---
NCCL1 1 MLQFAHG--------SAASGHGLGSNYN----APASARTWH---RMCPYGIPLTQPRAARTAGRL---GPSAT----------TSGVARGLA--SATSGPAAAVLSGGGPRLATTAPSATG AGG---
NCCL16 1 MLQVAQPRAARTAGCL---GPSAT----------AGGVARGLA--SATSGPAAAVLSGGGPLLAATAPAATG▼SGG---
NCCL30 1 MLAQL-SRALILASATGARGQALGSLSSAIRNRCLEHHQHLTTTTEDGC-VVRSAPLGSRCVGLS---HPAAAAGSEAPAAADHSSDLLQLPLPLPASHLAYTTSYSSVISGPAAAGLSGGGPRLAAAAAPAMF TPGAAA
NCCL34 1 MLQFARASA-------AASGHGLGSSYTS---APVSAHTGLT---LRPCGIPLTQPRAARTAGRL---QPLAT----------AGGVARGLACSSATSGPVAAVLSGGGPCLATAAPAVTG AGG---
NCCL33 1 MLQFARASA-------AASGHGLGSSYTS---APASAHTGLT---LRPCGIPLTQPRA---AGRL---HPPAT----------AGGVARGLACSSATSGPAAAVLSGGGPCLATAASAVTG AGG---
NCCL28 1 MLQFARASA-------AASGHGLGSSYTS---APASAHTGLT---LRPCGIPLTQPRASRTAGGL---HPSAT----------AGGVARGLACSSATSGPAAAVLSGCGPRLAAAAPAATG AGG---
NCCL29 1 MLQFARGSA-------AASRHGLGSSHTS---APVSARTGLT---LRPCGIPRTQPRA---AGRL---HPPAT----------AGGVARGLACSSATSGPAAAVLSGGGPCLATAASAVTG AGG---
NCCL27 1 MLQFARGSA-------AASGHGLGSSYTS---APVSAHTGLT---LRPCGIPLTQPRA---AGRL---HPSAT----------AGGVARGLACSSATSGPVAAVLSGGGPCLATAAPAVTG AGG---
NCCL31 1 MLQFARGSA-------AASRHGLGSSHTS---APVSAHTGLT---LRPCGIPLTQPRA---AGRL---HPSAT----------ARGVARGLACSSATSGPVAAVLSGGGPCLATAAPAVTG AGG---
NCCL32 1 MLQFARASA-------TASRHGLGSSHTS---APVSAHTGLT---LRPCGIPLTQPRA---AGRL---HPSAT----------AGGVALGLACSSATSGPAAAVLSGGGPCLATAAPAVTG AGG---
NCCL26 1 ML-TMR----------TRL------FGNTRNSHGCFTV S-----
NCCL3 1 M----RTGLL----------GR----PPL------SAQSNSGHGY--G Y-----
NCC2 1 M----RT----------SR----PPL------SAYSNSRHGY--G Y-----
NCCL5 1 MLAMRT----------SR----PPL------SAYSNSRHGY--G Y-----
NCCL6 1
NCCL13 1 MLVMRA----------S---SGLVAGSFSGTCARSHSGHGY--C N-----
NCCL10 1 MLALRV----------V---AGSPA------GAHRHSVHGC--S H-----
NCCL7 1 MLALRV----------V---AGSPA------GAHRHSVHGC--S Q-----
NCCL8 1
NCCL9 1 MLALRV----------V---AGSPA------GAHRHSVHGC--S Q-----
NCCL11 1 MLVMRA----------SGRFAGSPA------GAHSHLSGHG--S H-----
NCCL12 1
consensus 1 ml p a va ss sgp aa ls gp l t g

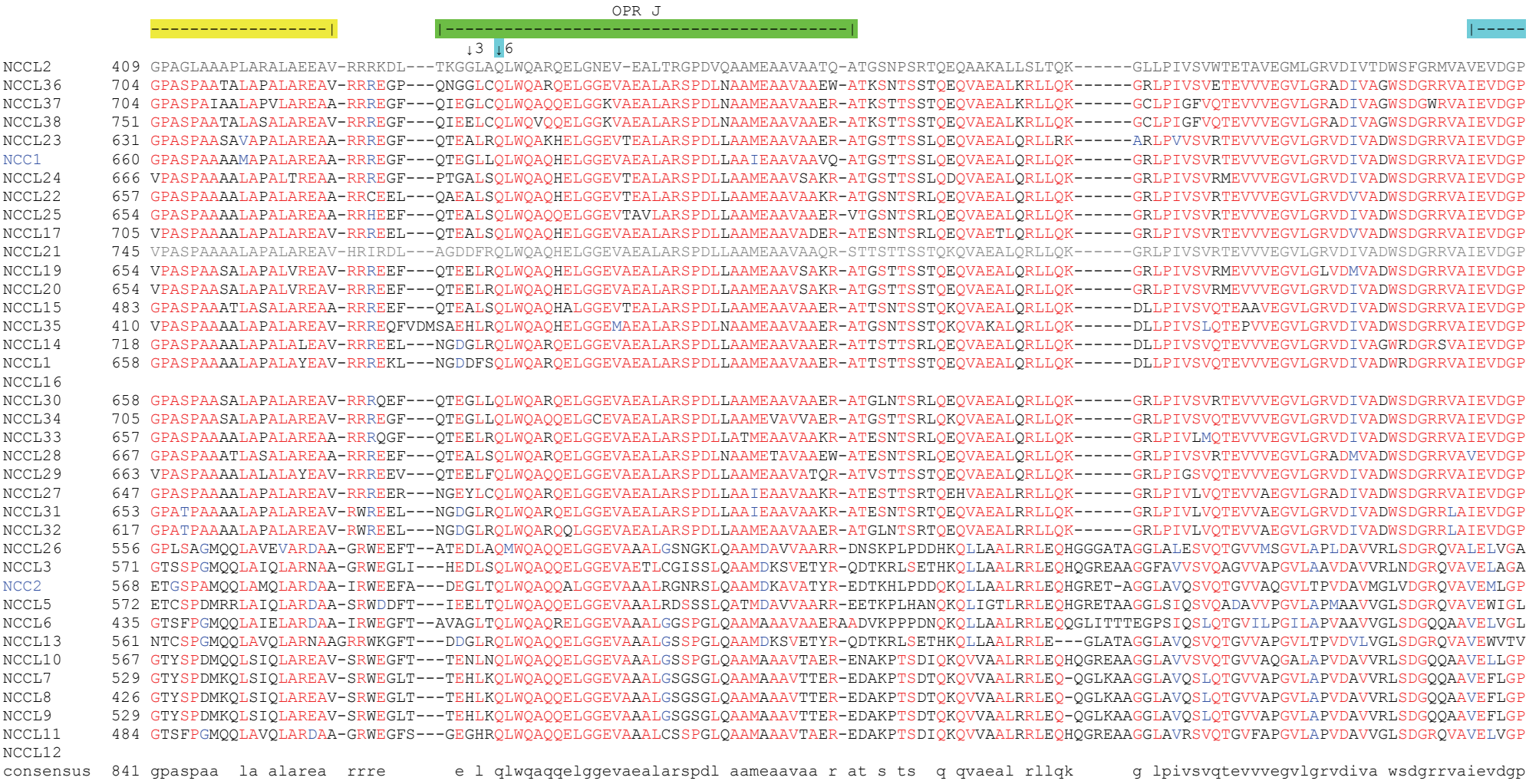

NCCL2 1
NCCL36 100 --GLRQLLPLAAWARQTAQAPATASPAAAGARSCLLQPGRRAQQQVLAAAGR---YGSDRSSDS---SS----------SSGRGYI------SGSRS----SSDGGGRVGRGRAGGGRWAGGGG-RGTGRTGGG-GG
NCCL37 98 --GLQQVLPLAAWARQAAPAPTAAAPAAAGARSHLLQPGRRAQQQVLAAAGR---YGSGGTSDSSSDSS----------SSRRGGI------SGSRS----SSDGGGRAGRGRADRGRWAGGGG-RGSGRTGGG-G-
NCCL38 98 --GLQQVLPLAAWARQAAPAPTAAAPAAAGARSYLLQPGRRAQQQVLAAAGR---YGSGGTSDRSSDSS----------SSRRGGI------SGSRS----SSDGGERAGRGRADGGRWAGGGG-RGSGRAGGG-G-
NCCL23 67 --GLRQLLPLAAWARQAAPAPATAAPAAAGARSCLLQPGRRAQQQVLAAAGR---YGSGSTSHRSIVSS-SDSSSSG-H-------GGI------GGSRS----SSDGGGRAGRGRAGRGHFASVGG-RGSGRTGGG--G
NCC1 94 --GLRQLLPLAAWARQAAPAPATAAPAAAGARSCLLQPGRRAQQQVLAAAGR---YGSGGTSHRSSGRS-SDNSSSG-H-------GGI------SGSRS----SSDGGGRAGRGRADGGHWAGGSG-RGSGRTGGG--G
NCCL24 100 --GLRQLLPLAAWARQAAPAPATAAAAAAGARSCLLQPGRRAQQQVLAAAGR---YGSGGTSDKSSGRS-SDNSSSG-H-------GGI------SGSRS----SRDGGGRAGRGRADGGRWAGGGG-RGSGRTGGG--G
NCCL22 100 --GLRQLLPLAAWARQAAPAPATAAPAAAGARSCLLQPGRRAQQQVLAAAGR---YGSGGTSDRSSGRS-SDNSSSG-R-------GGI------SGSRS----SRDGRGRAGRGRADGGRWAGGGG------------R
NCCL25 96 --GLRQLLPLAAWARQAAPAPATAAAAAAGARSCLLQPGRRAQQQVLAAGGR---YGSGGTSHRSIDR----SSSSG-H-------GGI------GSSRS----SSDAGGRAGGGRAGGGRWAGG-------RTGGA--G
NCCL17 97 --GLRQLLPLAAWARQAAPAPATAALAAAGARSCLLQPGRRAQQQVLAAAGR---YGSGSSSGRGSD-----NSSSG-H-------GGI------SGSRS----SSDGGGRAGRGRADGGRWAGGSG-RGSGRTGGG--G
NCCL21 94 --GLRQLLPLAAWARQAAPAPATAAPAAAGARSCLLQPGRRAQQQVLAAAGR---YGSGGTIDRSSGR----SSSSG-H-------GGI------GSSRS----SSDAGGRAGGGRAGGGRWAGG-------RTGGA--G
NCCL19 97 --GLRQLLPLAAWARQAAPAPAPAAPAAAGARSCLLQPGRRAQQQVLAAGGR---YGSGGTSHRSIDR----SSSSG-H-------GGI------GSSRS----SSDAGGRAGGGRAGGGRWAGG-------RTGGA--G
NCCL20 97 --GLRQLLPLAAWARQAAPAPAPAAPAAAGARSCLLQPGRRAQQQVLAAGGR---YGSGGTSHRSIDR----SSSSG-H-------GGI------GSSRS----SSDAGGRAGGGRAGGGRWAGG-------RTGGA--G
NCCL15 1
NCCL35 1
NCCL14 97 --GLRQLLPLAAWARQAAPAPATAAPAAAGARSCLLQPGRRAQQQVLAAAGR---YGSGGTSHRSSGRS-SDSTSDS-SIDSSSGRGSD------SSSNNDRSRGGHGADRAGRGRAGGGRWAGGRTGGGSGRTGGG---
NCCL1 95 --GLRQLLPLAAWARQAAPAPATAAPAAAGARSCLLQPGRRAQQQVLAAAWR---YGSGGTSYTSSDRS----------SD------SSNNNDKGRGGHVEDRAGRGRAGGGRWAGGRTGGGSGRSGGG-GG
NCCL16 61 --GLRQLLPLTAWARQAAAAPATAAPAAAGARSCLLQPGRRAQQQVLAAAGR---YGSGSTSDKSSGRS----------SG------SSNNNDKGRGGHVEDRTGRGRAGGGRWAGGRTGGGGSRS------
NCCL30 136 AAAYGSLPPPLGW----------HAQPQVAAAGVM---HGSGTTGSASSGSS-GSS-----S----------SS----SSSGEGDGGNGRSSGG----------GRGSGRTGG
NCCL34 99 --GLLQLLPLAAWARQAAPAPATAAPAAAGARSCLLQPGRRAQHQVLAAAGR---YGSGGAIDRSSGMS-SSS--------SSSGRDDI------SGSRS----SSDGGGRAGRGRAGGGRWNGGRRG---------SGR
NCCL33 96 --GLRQLLPLAAWARQAAPAPATGAPAAAGARSCRLQPGRRAQHQVLAAAGR---YGSGSSSGRG----------SDNSSSGHGGI------SGSRS----SSDGGGRAGRGWAGGGHWADGRTGGGGSMSAGD---
NCCL28 99 --GLRQLLPLAAWARQAAPAPAAVAPAAAGARSCLLQPGRRAQQQVLAAAGR---YGSGGIIDRSSGMS-SSSSSSSSSSSSSSGRDDI------SGSRS----SSDGGGKAGRGRAGGGRWNGGRRG---------SGR
NCCL29 96 --GLWQLLPLAAWARQAAPAPATAAPAAAGARSCLLQPGRRAQQQVLAAVGR---YGSGGIIDRSSGMS-SSSSSS--SSSSSSGRDDI------SGSRS----SSDGGGKAGRGRAGGGRWNGGRRG---------SGR
NCCL27 96 --GLRQLLPLAAWARQAAPAPATAAPAAAGARSCLLQPGRRAQQQVLVAVGR---YGSGGIIDRSSGRS-SDS----------SS----SSDGGGRAGRGRAGGGRWAGGRTGGAGGRSSG----
NCCL31 96 --GLRQLLPLAAWARQAAPAPATAAPAAAGARSCLLQPGRRAQHQVLVAAGR---YGSGGIIDRSSGMS-SSS----------RS----SSDGGERAGRGRAGGGRWARGRTGGAGGRSSGGDGR
NCCL32 96 --GLRQLLPLAAWARQAAPAPATGAPAAAGARSCRLQPGRRAQQQVLAAAGR---YGSGS----------RS----SSDGGGRAG----------RGR
NCCL26 23 ----------HMLAVSKGPRPAGALL----------PKTAG----------PSS----------SVPRPSQRAGSAVAVLGGGSEL--FLVPVASRQ---LVQAAAAGGGAAGGEGGWGRSGGLGGGHAGRDSS
NCCL3 24 ----------GYAAVSQGPRPASRVL----------LPVRSGTTVPRAFAGPSAARVAAPAPALR---S------SSTWSGSEAVLGGIRRLAPATPPTSRRQ---VVQTAAAGSGAAAGGANSGG-RGSGRG-YRGSGT
NCC2 21 ----------GSAAVSQGPRPASRGL----------LPVRSGTTVPRAFASPSSAYSARPVLLES---RSLSPR-SGSVPSVLAGSGGRRPIAPAVPPPSSRQ---LEQAAAGGGGASSS-----S-RGEGR--MGGRGR
NCCL5 24 ----------GYVAVSQGPRPASRVL----------LPVRSGTTVPRAFASPSS---ARPALLES---GSLSPRRSGSVPTVLAVGSGRRPIAPAVPPPSSRQ---LEQAAAGGGGAGSS-----S-RGEGR--MGGRGR
NCCL6 1
NCCL13 31 ----------SSSARLQAPQPACALR----------RLFGR------SDSDPSSPRAAAATSVPAHNSSSSSPL----RVSTAAVLGGS------SSSRLGRS---VVQVAAGRGSSGGRAGGG--------------S-
NCCL10 25 ----------SLSARLQAPPPACAFR----------RLFVYGMHRTAADTGPSSPRAAAATSVPALNSSSSSPL----RVSTAVVLCGS------SSSRLGRP---IVQAAAGRGSGGGRAGDG--------------SR
NCCL7 25 ----------SLSARLQAPPPACAFR----------RLFVYGMHRTAADTGPSSPRAAAATSVPALNSSSSSPL----RVSTAVVLCGS------SSSRLGRP---IVQAAAGRGSGGGRAGDG--------------SR
NCCL8 1 ----------GRGSGGGRAGDG--------------SR
NCCL9 25 ----------SLSARLQAPPPACAFR----------RLFVYGMHRTAADTGPSSPRAAAATSVPALNSSSSSPL----RVSTAVVLCGS------SSSRLGRP---IVQAAAGRGSGGGRAGDG--------------SR
NCCL11 28 ----------RSPACLQAPRPVVTAM----------PAARGIAPSRAPASDPSSARAAATASLRARY--SPFSL----RGLMVAMRGGS------SSTRLARP---VVQAAAGRGSGGGRGRGGGMAGG------GRGG
NCCL12 1 MPGRL----QATLTATRGFL------AQPQVQRR---FVQTAGGRGSGGGGGRTGG------------GG
consensus 141 g l p aawar ap pa aa pgrraq qv aaag ygsg rs g agrg aggg gg

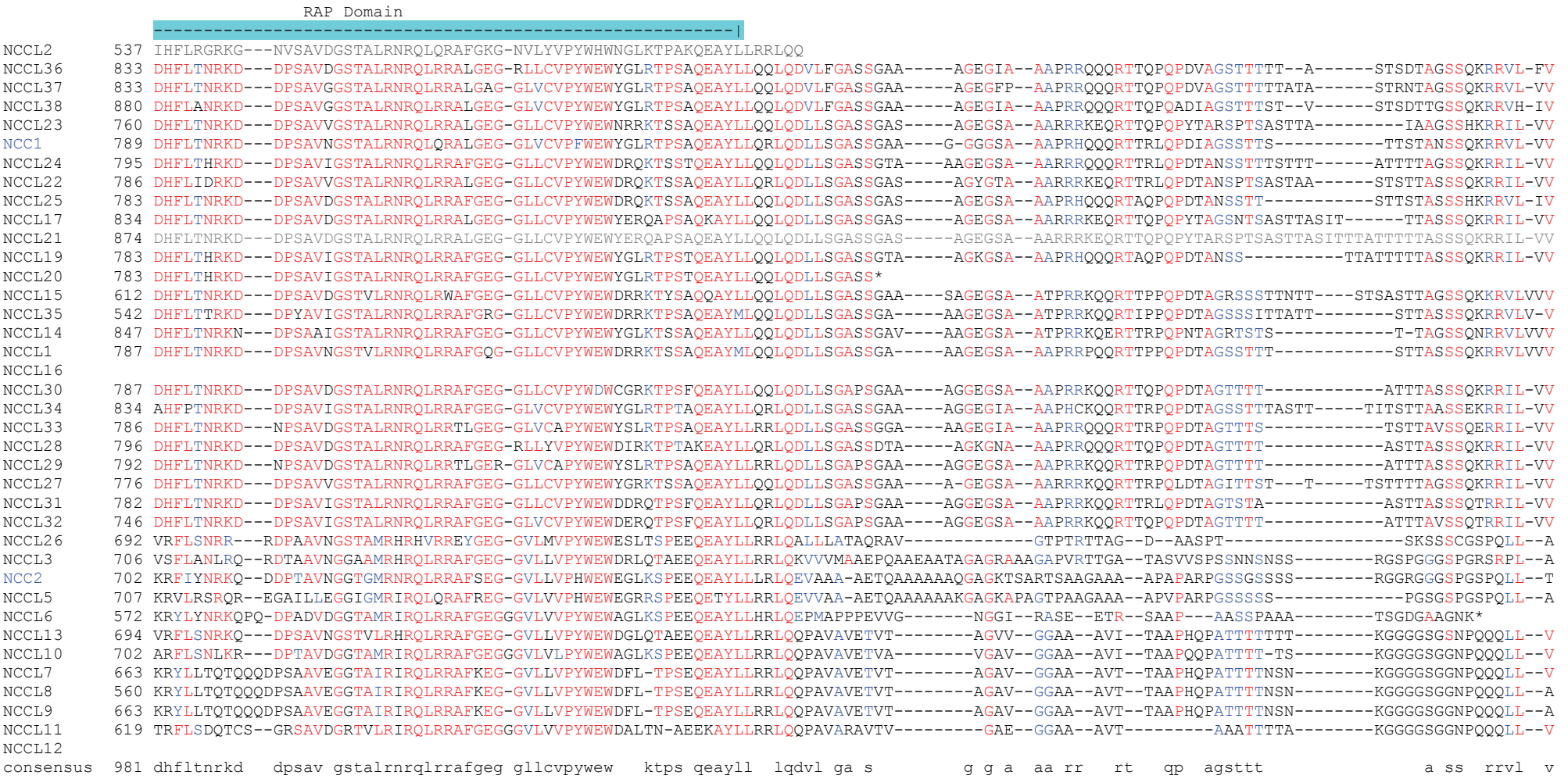

OPR A
NCCL2 1 MSLL-APSRPAARVSTRRKAATAAIRASIIATLTPALLPLVPRIRNPFGCSIPLWVLAKAGA---ASDERVESQLAPVLLQRLGD-PVLLKHAAPQ
NCCL36 207 --GGGGGRARDPADDMTMCRTLEALQAVIDRRLAVWSERQDVSTMSAAFGRCGK▼LESACAGGPAATAAARAGIMAALAPALLPLVPRIRQPVGCSLPLWALARAGA---ASDGHVESQLAPALLQRLVD-PVLLEAAKPQ
NCCL37 207 --GGGGGRPRDPSDDMTMCRNLEALQAVIGQRLSMWSGRQDVSTMSVAFNLCGK▼LESACAGGPAATAAARAGIMAALACAYLPLVPRIRHPADCSIPLLALAKTGA---ASDGRVESQLAPALLQRLVD-PVLLEAAKPQ
NCCL38 207 --GGGGGRARDPADDMTMCRNLEALQAVIGQRLSVWSERQDVSTMSVAFNLCGK▼LESACAGGPAATAAARAGIMAALACAYLPLVPRIRHPADCSIPLLALAKTGA---ASDGRVESQLAPALLQRLVD-PVLLEAAKPQ
NCCL23 180 ---GGGGGRARDPADDMTVCRTLEELQGVIDQRLAVWCERQDVSSMSAAFNLCGK▼LESARAGGPAATAAARAGIIAALAPALLPLVPRIRQPAGCSLPLLALSKAGA---ASEGRVESQLASALLQRLVD-PVLLDSATPQ
NCC1 207 GRSGGGGRPRDPEDVMIGYATLEELQGVIDQRLAVWCERQDVSTMSAAFGRCGK▼LESARAGGPTATAAARANIIAALAPALLPLVPRIRQPAGCSLPLLALAKAGA---ASDGRVESQLAPALLQRLVD-PVLLDSATPQ
NCCL24 213 GRSGGGGRPRDPEDVMIGYATLEELQGVIDQRLAVWCERQDVSTMSAAFGRCGK▼LESARAGGPAATAAARAGIIAALAPALLPLVPRIRQPAGCSIPLLALSKAGA---ASDGRVESQLAPALLQRLVD-PVLLDSATPQ
NCCL22 204 GRSGGGGRPRDPEDLMIGYATLEELQGVIDQRLAVWCERQDVSTMSAAFGRCGK▼LESARAGGPTATAAARAGIIAALAPALLPLVPRIRKPVDCSIPLLALAKAGA---ASDGRVESQLAPALLQRLVD-PVLLDGATPQ
NCCL25 200 GRSGGGGRARDPADDMTVCRTLEDLQAVLDQRLAVWCERQDVSTMSAAFGRCGK▼LEPARAGGPAATAAARAGIIAALAPALLPLVSRIRQPAGCSIPLWALAKAGA---AIDGRVESELAPALLQRLVD-PVLLDSAIPI
NCCL17 206 GRSGGGGRPRDPEDVMIGYATLEELQGVIDQRLAVWCERQDVSAMSAAFGRCGK▼LESARAGGPTATAAARAGIIAALAPALLPLVPRIRQPASCSIPLLALSKAGA---ASDGRVESELAPALLQRLVD-PVLLESATPQ
NCCL21 198 DRSGGGRRARDPADDMTVCRTLEQLQGVIDQRLAVWCERQDVIAMSAAFGRCGK▼LESARAGGPAATAAARAGIIAALAPALLPLVPRIRQPAGCSLPLWALAKAGAA*AASDGRVESQLAPALLQRLVD-PVLLDSATPQ
NCCL19 201 DRSGGGRRARDPADDMTVCRTLEQLQGVIDQRLAVWCERQDVIAMSAAFGRCGK▼LEPARAGGPTATAAARAGIIAALAPALLPLVPRIRQPASCSLPLWALAKAGA---ASDGRVESQLAPALLQRLVD-PVLLDSAIPI
NCCL20 201 DRSGGGRRARDPADDMTVCRTLEQLQGVIDQRLAVWCERQDVIAMSAAFGRCGK▼LESAHASGPAATAAARAGIIAALAPALLPLVPRIRQPAGCSLPLWALAKAGA---AIDGRVESELAPALLQRLVD-PVLLDSATPQ
NCCL15 1 MIGYATLEALQAVIDRRLAVWSERQDVSTMSAAFGRCGK▼--SARAGGPTATAAARAGIMAALAPALLPLVPRIRQPADCSLPLWALAKAGA---ASDGRVESQLAPALLQRLVD-PVLLKTATPQ
NCCL35 1 MSAAFNLCGK▼LESARAGGPAATAAARAGIMAALAPALLPLVPRIRQPAGCSIPLWALAKAGA---ASDGRVASQLAPALLQRLLD-PVLLEAAKPQ
NCCL14 221 --GGGGGRARDPADDMTVCRTLEDLHDVIDKRLAVWSERQDVMTMSAAFNLCGK▼LESARGGGPAATAAARAAIVAALAPALLPLVPRIRQPAGCSLPLWALAKAGA---ASDGRVESQLAPALLQRLVD-PVLLKAAKPQ
NCCL1 205 GGGGGGGRARDPADDMTVCRTLEDLQAVIDRRLAVWSERQDVSTMSAAFNLCGK▼LESARTSGPAPTAAARFRIVAALSPALLPLVPRIRQPAGCSLPLWALAKAGA---AIGGRVESQLAPALLQRLAD-PVLLKTATPQ
NCCL16 166 --GGGGGKARDPADDMSVCRTLEDLHDVIDRRLVVWSERQDVMTMSAAFNLCGK▼LDSARAGGPAATAAARAGIVAALAPALLPLVPRIRQPAGCSIPLLALAKAGA---AIDGRVESQLAPALLQRLAD-PVLLKTATPQ
NCCL30 206 NGGGGGGRARDPTDDMTVCRTLEDLQAVIDQRLAVWCERQDVSAMSAALGRCGK▼LESARA-GPAATAAARAGIIAALAPALLPLVPRISKPVDCSIPLWALAKAGA---ASDGRVESQLAPALLQRLVD-PVLLDSATPQ
NCCL34 206 SSGGGGGRARDSADDMTMCRTLEELQAVIDQRLAVWCERQDVSTMSAAFNLCGK▼LESSHAGGPATTAAARAGIMAALAPALLPLVPRIRQPASCSIPLWALAKAGA---ASDGRVESQLAPALLQRLVD-PVLLDSATPQ
NCCL33 205 -RSGGGGRARDPADDMTVCRTLEQLQGVIDQRLAVWCERQDVSTMSAAFNLCGK▼LESAHAGGPAATAAARAGIIAALAPALLPLVPRIRQPASCSLPLLALAKAGA---ASDGRVESQLASALLQRLVD-PVLLETATPQ
NCCL28 214 SSGGGGGRARDSADDMTMCRTLEELQAIIDQRLAVWSVRQDVSTMSAAFNLCGK▼LESARAGGPAATAAARAGIMAALAPALLPLVPRIRQPAGCSIPLWALAKAGA---ASDGRVESQLAPALLQRLVD-PVLLKTATPQ
NCCL29 209 SSGGGGGRARDSADDMTMCRTLEELQAISDQRLAVWSVRQDVSTMSAAFNLCGK▼LESARAGGPAATAAARAGIMAALAPALLPLVPRIRQPAGCSIPLWALAKAGA---ASDGRVESQLAPALLQRLVD-PVLLDSATPQ
NCCL27 197 ---AGGVRARDPADDMTVCRTLEDLQAVIDQRLAVWCERQDVMTMSAAFNLCGK▼LESARAGGPAATAAARAGIMATLAPALLPLVPRIRNPASCSLPLWALAKAGA---ASDGRVESQVAPALLQRLVD-PVLLDSATPQ
NCCL31 201 TGGGGGGRARDPADDMTVCRTLEELPAVIDQRLAVWSERQDVMTMSAAFNLCGK▼LESSHA-GPAATAAARAGISAALAPALLPLVPRIRQPASCSIPLWALAKAGA---ASDGRVESQLAPALLQRLVN-PVLLDSATPQ
NCCL32 165 AGGGGGGRARDPADDMTVCRTLDELQVVIDQRLAVWCERQDVMTMSAAFNLCGK▼LESSHA-GPAATAAARAGIMAALAPALLPLVPRIRQPASCSIPLWALAKAGA---ASDGRVESQLAPALLQRLVN-PVLLGMAKPQ
NCCL26 112 GRGRGRGQGQDST---PNDGTLEELQAAVTDQLPGWVERHDTAAISNAFRKAVQ▼LRSG-DGGAATAARIRSGIIAGLAAGYLRLVPRIRKAIECRIPLWAMAKAGIS--GK--GKEAELTAALLQRLVTDSRLLGAATPM
NCCL3 130 GRGGGSRGRGSPP---ADNQTLEELRAVVNGQLPGWVARKDNASIWRAFNKAAK▼LRSG-DGGAATAARIRSGIIADLSAAYLPLGPAVCAAVECSIPLWAMAKVGIS--GQ--GKEAQLAAALLERLLD-PAVLSTATPM
NCC2 126 GRGGFRGRGGTAA---SDDETLEELQAAVTCQLPGWVEREDTAAISNAFRKAVQ▼YRSG-NGGAATAARIRSGIIADLSAAYLPLVPRIRKPFDCRMPLWALGKAGIG--SK--GPEVQLANALLQRLLD-PAVIAAATAM
NCCL5 127 GRGGFRGRGGTAA---SDDEMLEELQAAVTCQLPGWVEREDTAAISNAFRKAVQ▼LRSA-DGGAAMAARISSGIIADLSAAYLPLVPRIRKAIDCRMPLWALGKAGIG--SK--GPEAQLANALLQRLLD-PAVIAAATPM
NCCL6 1 MTVCQTLEELQALISQRLSVWKGRKDVTTMFAAFNLCGR▼LKSR-AG-AAATAVARNSIIGALAPAYLPLVPRIRDAKNCSIPLLALGKAGVA--GGGGGTEAQLAAALLERLVD-PEVLVDAQPQ
NCCL13 117 ASGGGRGSGRDPADDMSVCQNLEDLQTFITRRLPVWEGKGDVITMCAAFNKCGK▼LESR-AG-ANQTVVARSSIIAALAPAYLPLVPRMRMAVDCSLPLWALGKAGVS--GD---TEAQLAAALLERLVV-PEVMGTAKPQ
NCCL10 118 ASGRSRGSGRDPADDMTVCRTLEDLQAIISQRLSVWEGRKDFITMCAAFNLCGK▼LESR-AG-ASQTAVARSSIIGMLAPAYLPLVPRIRKAVDCSIVLWALGKAGIS--GG-SAMEAQLIAALLERLLD-SEVLGTAKPQ
NCCL7 118 ASGRSRGSGRDPADDMTVCRTLEELQAIISQRLSVWEGRKDFITMCAAFNLCGK▼LESR-AG-ASQTAVARSSIIGMLAPAYLPLVPRIRNAKNCSIPLWALGKAGLV--GG-G-TEAQLAAALLERLVD-PEVLGKAEPQ
NCCL8 15 ASGRSRGSGRDPADDMTVCRTLEELQAIISQRLSVWEGRKDFITMCAAFNLCGK▼LESR-AG-ASQTAVARSSIIGMLAPAYLPLVPRIRNAKNCSIPLWALGKAGLV--GG-G-TEAQLAAALLERLVD-PEVLGKAEPQ
NCCL9 118 ASGRSRGSGRDPADDMTVCRTLEELQAIISQRLSVWEGRKDFITMCAAFNLCGK▼LESR-AG-ASQTAVARSSIIGMLAPAYLPLVPRIRNAKNCSIPLWALGKAGLV--GG-G-TEAQLAAALLERLVD-PEVLGKAEPQ
NCCL11 127 SGGSGRGSGRDPADDMTVCRTLEELQATISQRLPVWEGRKDFITMCAAFNLCGR▼LESS-GGPAAATAAARSSIIGTLAPAYLPLVPRIRKARDCSIPLWALGKAGLF---G-GGTEAQLAAALLERLVD-PEVLAKAKPQ
NCCL12 47 ASGGGRGGARDPADDMDVCQNLEDLQALISQRLSVWERKKDVVTMCAAFSKCGN▼LESR-SAGAAATAAARTGILAVLAPAYLPLVPRIRNAMKCSIPLWALGKAGVS--SS-GGTEAQLVAALLERLLD-SEVLGTARPQ
consensus 281 ggggggrardpaddmtmcrtleelqavi qrl vw erqdv tmsaaf cgk lesa aggpaataaaragIiaaLapalLpLvprirnp CsipLwalakaGa asdgrvesqlapaLLqRLvd pvll A pq


OPR B
OPR C
OPR D


OPR E...
...OPR E
↓6
↓3 ↓6
NCCL2 227 ELTNILMALQGLP----------QIGRQQSELLAAAVAAVDLQRDFAGYNSQDLSDSAW
NCCL36 477 HLSNILLALEGL----------QLGGKQSELLAAAVAAECVRRLRTLKEAEAFIPQHLSNILLALEGLQLRGKQSELLASAVATEGVRRGFARFESQALSNSAW
NCCL37 477 ELSNILLALEGL----------QLGGTQSELLAAAVAAECVRRLRTLKEAQAFKPQGLSNILLALEGLQLRGKQSELLASAVAAEGMRRAFAGFKPQDFSNAAW
NCCL38 477 ELSNILLALEGLQLGGKQSELLAAAVVAECVRRLRTPKEAEAFKPQELSNILLALEGLQLGGKQSELLAAAVAAECVRRLRTLKEAEAFIPQELSNILLALEGLQLRGKQSELLASALAAEGMRRAFAGFKPQDFSNSAW
NCCL23 450 GLSNILLALEGL----------QLRGKQSELLTAAVAAEGVRRGFAGFKPQELSNSAW
NCC1 479 HLSNILLALEGL----------QLRGKQSELLASAVAAEGVRRGFAGFNPQDLSNSAW
NCCL24 485 GLSNILLALEGL----------QLRGMQAELLAAAVAAEGVRRGFAGFEPQHLSNSAW
NCCL22 476 GLSNILLALEGL----------QLRGKQSELLASAVAAEGVRRGFAGFEPQHLSNSAW
NCCL25 473 GLSNILLALEGL----------QLRGMQAELLAAAVAAEGMRRGFAGFKPQELSNSAW
NCCL17 478 HLSNILLALEGL----------QLGGKQSELLAAAVAAECVRRLRAPKEAEAFIPQHLSNILLALQGLQLRGKPSELLADAVAAEGVQRGFAGFKPQELSNSAW
NCCL21 472 DLSNILLALEGLQLGGKQSELLAAAVAAECVRRLRAPKEAEAFKPQELSNILLALEGLQLGGKQSELLAAAVAAECVRRLRAPKEAEAFKPQELSNILLALEGLQLGGKQSERLAAAVAAEGVRRGFAGFKPQELSNSAW
NCCL19 473 HLSNILLALEGL----------QLGGKQSERLAAAVAVEAMRRGFAGCNPQELSNSAW
NCCL20 473 HLSNILLALEGL----------QLGGKQSERLAAAVAVEAMRRGFAGCNPQELSNSAW
NCCL15 256 HLSNILLALEGL----------QLRGMQAELLTAAVAAECVRRLGTPKEATAFIPQELSNILLALEGLKLRGKQSELLAAAVAAEGVQRGFAGFNPQCLSNSAW
NCCL35 229 GLSNILLALEGL----------QLGGQQSERLAAAVAAEGVQRGFAGFNPQALSNSAW
NCCL14 491 HLSNILLAMEGL----------QLRGMQAELLAAAVAAECVRRLRTPMGAEAFKPQELSNILLALEGLQLRGKQSELLAAAVAAEGVRRGFAGFNPQCLSNSAW
NCCL1 477 DLSNILLALEGL----------QLRGKQSELLASAVAAACVRRGFAGFEPQHLSNSAW
NCCL16
NCCL30 477 GLSNILLALEGL----------KLRGKQSRLLAAAVAAECVRRGFAGFKPQHLSNSAW
NCCL34 478 GLSNILLALEGL----------QLGGTQSERLAAAVAAECDRRLRARKEAEVFKPQELSNILLALEGLQLGGKQSERLAAAVAAEGVRRGFAGFEPQHLSNTAW
NCCL33 476 HLSNILLALEGL----------QLGGKQSELLAAAVAAEGVRRNFAGFNPQALSNSAW
NCCL28 486 ELSNILLALEGL----------QLGGKQSERLAAAVAAEGVWRGFAGFTPQDLSNSAW
NCCL29 482 ELSNILLALEGL----------QLGSKQSERLAAAVAAEGVWRGFAGFTPQELSNSAW
NCCL27 466 ELSNILLALEGL----------QLGGKQSELLAAAVAAEAMRRGFAGFNPQHLSNSAW
NCCL31 472 GLSNILLALEGL----------QLGGTQSEQLAAAAVAVEAMRRGFAGCNPQELSNSAW
NCCL32 436 GLSNILLALEGL----------QLGGTQSEQLAAAAVAVEAMRRGFAGCNPQELSNSAW
NCCL26 376 HLSNILLALEGL----------QLGGKQAELL-AAVATEDLRRGFSGYNDQDLTNSAW
NCCL3 393 HLSNILLALEGL----------QLGREKAQLV-AAVAADGVRRGFAGFKPQELSNSAW
NCC2 389 DLSNILLALEGL----------QLGSEQAQLV-SAVAAEDVRRGFTRYNSQDISNSTW
NCCL5 395 HLSNILLALEGL----------QLGSDQAQLTV-AVAAEDVRRSLAGYVAQDISNSAW
NCCL6 259 DLSTILLALEGL----------QLGGKQAELV-AAVAAEDVRRGFAGYGAQNVGNSAW
NCCL13 385 HLSNILLALEGL----------QLGGKQAELV-AAVAAEDVRRGFDGYVAQDLSNSAW
NCCL10 388 HLSNILLALEGL----------QLGGEQAELVV-AVAAEDVRRGFAGYVAQDLSNSAW
NCCL7 350 QLSNILLALEGL----------QLCSEQAELV-VAVAEEDMRRGFDGYIAQDLSNSAW
NCCL8 247 QLSNILLALEGL----------QLCSEQAELVV-AVAAEDVRRGFDGYIAQDLSNSAW
NCCL9 350 QLSNILLALEGL----------QLCSEQAELVV-AVAEEDMRRGFDGYIAQDLSNSAW
NCCL11 306 ELSNVLLALEGL----------QLGGGQAGLVV-AVAAEDVRRGFTGYVPQDLNNSAW
NCCL12 227 ELSNILLALEGL----------QLGGKQAELV-AAVAAEDVRRGFEGYVPQDLSNSGR
consensus 561 lsnillalegl ql g qse laaavaaegvrrgfagy pqdlsnsaw


```
NCCL2      276 ALAKMGYGQGTTPLSSKQSQWYAAAVAAAQR---PGVMASAKPQAWANLLYALALVRYQPSPTLLEAGAASVMES---GTAHSCAITLWALAVLQLRHAGVESAVCDRLGELLWLHPASVPVQHLSSSLWALAVLAGG-G
NCCL36     571 ALAKMGYGAGSPSEAMEQRQWYAAAVAAAQR---PGVMASAKPQAWANLLYALALVRHQPPPVLLDAGAAAAMQR---GNGQECANTLWALAVMRLRHAGLEAAVCGRLGELLRLGPESLAVQELCNSLWALAVLAGG-G
NCCL37     571 ALAKMGYGAGSPSQAMEQRQWYAAAVAAAQR---PGVMASAKPQAWANLLYALALVRHQPPPALLDAGAAAAMQR---GSAQECANTLWALAVMQLRHAGLEAAVCGRLGELLRQDPESLVVQNLCNSLWALAVLAGG-G
NCCL38     617 ALAKMGYGAGSPSEAMEQRQWYAAAVAAAQR---PGVMGSAKPQAWANLLYALALVRHQPPPALLDAGAAAAMQR---GSAQACANTLWALAVLQLRHAGLEAAVCGRLGELLRQDPELLAAQALCNSLWALAVLAGGGG
NCCL23     498 ALAKMGYGAVATPQATEQSHWYAAAVAAAQR---PGVMAGATPQNWANLLYALALLRHQPPPALLDAGAAAAMQR---GNAQECANALWALAVLQLRHAGLEAAVCGRLGELLQREPESLIAQELCNSLWALAVLAGG-D
NCC1       527 ALAKMGYGAGATPQAMEQSHWYAAAVAAAQR---PGVMAGAKPQDWANLLYALALVRHQPPPTLLDAGAAAAMQG---GNAQECSNTLWALALLQLRHAGLEAAVCGRLGELLRQGPESLIAQELCNSLWALAVLAGG-G
NCCL24     533 ALAKMGYGAGATPQATEHSHWYAAAVAAAQR---PGVMAGATPQNWANLLYALALVRHQPPPALLDAGAAAAMQR---GNAQHCANALWALAVLQLRHAGLEAAVCGRLGELLRQGPESLAEQALCNSLWALAVLSGG-G
NCCL22     524 ALAKMGYGVGATPQAMEQSPWYAAAVAAAQR---PGVMAGAKPQDWANLLYALALVRHQPPPVLLDASAAAAMQR---GSAQHCANSLWALAVLQLRHVGLEAAVCGRLGELLRQGPESLAEQDLCNGLWALVALAGG-G
NCCL25     521 ALAKMGYGAGATPQAMKQCPWYAAAAAAAQR---PGVMAGATSQNWANLLYALALVRHQPPPALLDAGAAAAMQR---GNAQDCANTLWALAVLQLRHAGLEAAVCGRLGELLRQEPESLAEQALCNSLWALAVLAGG-G
NCCL17     572 ALAKMGYGAGATPQAIEQSHWYAAAVAAAQL---PGVMAGAKPQAWANLLYALALVRHQPPPTLLDAGAAAAMQG---GNAQACANTLWALAVLQLCHAGLEAAVCGRLGELLRLGPESLVAQALCNSLWALAALAGG-G
NCCL21     612 ALAKMGYGAGATPQAMEQSHWYAAAVAAAER---PGVMAGAKPQAWANLLYALALVRHQPPPTLLDAGAAAAMQR---GTAQHCANTLWALAVLQLRHAGLEAAVCGRLGELLRLGPESLVEQALCNSLWALAALAGG-G
NCCL19     521 ALAKMGYGAGATPKAMEQCPWYAAAVAAAQR---PGVMASAKPQHWANLLYALALVRHQPPPALLDAGAAAAMQR---GKAQECANTLWALAVLQLRNAGLEAAVCGRLGELLQREPESLVAQALCNSLWALAVLAGG-G
NCCL20     521 ALAKMGYGAGATPKAMEQCPWYAAAVAAAQR---PGVMASAKPQAWANLLYALALVRHQPPPALLDAGAAAAMQR---GKAQECANTLWALAVLQLRNAGLEAAVCGRLGELLQREPESLVAQALCNSLWALAVLAGG-G
NCCL15     350 ALAKMGYGASATSQATQQSHWYAAAAAAAQR---PGVMAGAMPQNWANLLYALALVRHQPPPALLDAGAAAVMQR---GNAQDCANTLWALAVLQLRHAGLEAAVCGRLGELLRLEPESLVAQELCNSLWALAVLAGG-G
NCCL35     277 ALAKMGYGAGATPQATEQSHWYAAAVAAAER---PRVMAGAKPQEWANLLYALALVRHQPPPALLDAGAVVAMQR---GNAQECANTLWALAVLQLRHAGLEAAVCGRLGELLRLGPESLVAQNLCNSVWALAVLAGG-G
NCCL14     585 ALAKMGYGAGATSQATEQSQWYAAAAAAAER---PGVMAGAKPQEWANLLYALASVRHQPPPALLDAGAAAVMQR---GNAQECANTLWALAVLQLRHAGLEAAVCGRLGELLRLGPESLVAQELCNSLWALAVLAGG-G
NCCL1      525 ALAKMGYGAGATSQATQQSHWYAAAAAAQR---PGVMAGAKPQGWANLLYALALVRHQPPPALLDAGAAAVMQR---GNAQECANTLWALAVLQLRHAGLEAAVCGRLGELLRLEPESLIAQELCNSLWALAVLAGG-V
NCCL16
NCCL30     525 ALAKMGYGAGATLHAAEQRHWYAAAVAAAQR---PGVMAGAKPQAWANLLYALALVRHQPPPALLDAGAAAAMQR---GNAQACANTLWALAVLQLRHAGLEAAVCGRLGELLRQGPESLVAQDLCNSLWAVAVLAGG-G
NCCL34     572 ALAKMGYGAGATPQAIEQRPWYAATVAAAQR---PGVMAGAKPQDWANLLYALALVRHQPPPALLDAGAAAAMQR---GNAQHCANTLWAQAVLQLRHAGLEAAVCGRLGELLRLDPESLVAQALCNSLWALAVLAGG-G
NCCL33     524 ALAKMGYGAGATPKAMEQSPWYAAAVAAAQR---PGVMASAKPQYWANLLYALALVRHQPPPALLDAGAAAAMQR---GNVQECANTLWALAVLQLRHAGLEAAVCGRLGELLRQGPESLTAQELCNSLWALAVLAGG-G
NCCL28     534 ALAKMGYGAGATSQAIEQCPWYAAAVAAAQR---PGVMASAKPQAWANLLYALALVRHQPPPALLDAGAAVAMHR---GNAQECANTLWALAVLQLRHAGLEAAVCGRLGELLRQGPESLIAQELCNSLWALAVLAGG-G
NCCL29     530 ALAKMGYGADATSQAIEQSPWYAAVAAAAQR---PGVMAGAKPQAWANLLYALALVRHQPPPALLDAGAVAAMQR---GNAQECANTLWALAVLQLRHAGLEAAVCGRLGELLRQGPESLIAQELCNSLWALAVLAGG-G
NCCL27     514 ALAKMGYGAGATSQAMKQSPWYAAVTAAAQR---PGVMAGAKPQAWANLLYALALVRHQPPPALLDAGAAAAMQR---GDAQACANTLWALAVLHLCHAGLEAAVCGRLGELLRQGPESLIAQELCNSLWALAVLAGG-D
NCCL31     520 ALAKMGYGAGATSQAMKQSPWYAAVAAAAQR---PGVMAGAKPQAWANLLYALALVRHQPPPALLDAGAVAAMQR---GNAQDCANTLWALAVLQLRHAGLEAAVCGRLGVLLRLEPESLAAQNLCNSLWALAVLAGG-G
NCCL32     484 ALAKMGYGAGATSQAMKQSPWYAAVAAAAQR---PGVMAGAKPQAWANLLYALALVRHQPPPALLDAGAVAAMQR---GNAQDCANTLWALAVLQLRHAGLEAAVCGRLGELLRLGPESLAEQNLCNSLWALAVLAGG-G
NCCL26     423 ALAKMEFGVGPE-APEEQRKWISAAVDAAKRRGRDGVMAGATPQAWSNLLYALALVRHQPPPALLDAGAAAAIL--KSGNGQDCASVLYALAVLRLRHGELEAAVCGRLGALLKRDKQAVNHQCIATSLWGIAVP----G
NCCL3      440 ALAKMGFGAGPE-APEEQRQWVSAAVDAAMR---PGTMATATPQAWANLLYALALMRHQPPPALLQNAGAAAAML-SSGSGQNCANTLYALAVLQLRHAGLEAAVCGRLRELLRGDRQALNNQEVSNTLWAMAVL----G
NCC2       436 ALAKMGFGVGPE-APAEQRQWVTAALDAAMR---PGTMATATPQNWSNLLYALSVMRHQPPPVLLDAGAAAAMRANSVNGPQDCANTLYALAVLQLRHAGLEAAVCGRLGELQQEDLESLTEQGLANSLWAVAVF----G
NCCL5      442 ALAKMGFGVGPE-APEEQRQWVTAALDAAMR---PGTMATATPQAWANLLYALALMRHQPPPVLLDAGAA-AMRANGVNG-QECANTLYALAVLQLRHAGLEAAVCGRLGELLWGGKQALNHQDIANSLWAMAVF----G
NCCL6      306 ALAKMGLGVGPE-APTEQRQWVSAAVDAAMR---PGVMARATPQAWANLLYALALTRHQPPPGLLDAGAGESMQR---GNGQDCANTLWALAVLQLRHAGLEAAVVGRLGELLQHDSQALNQQDISSSLWAMAVL----G
NCCL13     432 ALAKMGFGVGPE-APVEQRQWVSAAVDAAMR---PGVMARATPQAWANLLYALSMIRHQPSPALLDAGAAAAMHR---GNAQDCANTLWALAVLQLRHAGLEAAVVGRLGELLQRGGEALTEQGISNSLWAMAAF----G
NCCL10     435 ALAKIGFGVGPE-APAEQRQWASAAVVAAMR---PGVMARATPQAWANLLYALALMRHQPPPELLDAGAAAAAMRASSAGGQNCANTLWALAVLQLRHAGLEAAVCGRLGELLQRDRQAVNNQNISNSLWAMAVI----G
NCCL7      397 ALAKMGFGAGPE-APAEQRQWVSAAVDAATQ---SGAMARATSQAWSNLLYALALMRHQPPAELLDSGAAVAAMRASSAGGQNCANTLYALAVLQLRHAGLEAAVCGRLGELLQHDSDTVLHQHVCNSLWAMAVF----G
NCCL8      294 ALAKMGFGAGPE-APAEQRQWVSAAVDAATQ---SGAMARATSQAWSNLLYALALMRHQPPAELLDSGAAVAAMRASSAGGQNCANTLYALAVLQLRHAGLEAAVCGRLGELLQHDSDTVLHQHVCNSLWAMAVF----G
NCCL9      397 ALAKMGFGAGPE-APAEQRQWVSAAVDAATQ---SGAMARATSQAWSNLLYALALMRHQPPAELLDSGAAVAAMRASSAGGQNCANTLYALAVLQLRHAGLEAAVCGRLGELLQHDSDTVLHQHVCNSLWAMAVF----G
NCCL11     353 ALAKMGFGVGPE-APAEQRQWVSAAVDAAMR---PGVMTGATPQEWANLLYALALMRHQPPPALLDAGAAA-AMRASSADGQDCANTLYALAVLQLRHAGVEAAVCGRLGELLQRGGKALTEQGISNSLWAMAVL----G
NCCL12     274 VALY*
consensus  701 alakmgygag    a eq  wyaaavaaa r   pgvma a pqawanllyalalvrhqpppalldagaaaamqr   g aq cantlwalavlqlrhagleaavcgrlgellr  pesl  q lcnslwalavlagg g
```


OPR J
↓3 ↓6
NCCL2 409 GPAGLAAAPLARALAEEAV-RRRKDL---TKGGLAQLWQARQELGNEV-EALTRGPDVQAAMEAAVAATQ-ATGSNPSRTQEQAAKALLSLTQK------GLLPIVSVWTETAVEGMLGRVDIVTDWSFGRMVAVEVDGP
NCCL36 704 GPASPAATALAPALAREAV-RRREGP---QNGGLCQLWQARQELGGEVAEALARSPDLNAAMEAAVAAEW-ATKSNTSSTQEQVAEALKRLLQK------GRLPIVSVETEVVVEGVLGRADIVAGWSDGRRVAIEVDGP
NCCL37 704 GPASPAIAALAPVLAREAA-RRREGF---QIEGLCQLWQAQQELGGKVAEALARSPDLNAAMEAAVAAER-ATKSTTSSTQEQVAEALKRLLQK------GCLPIGFVQTEVVVEGVLGRADIVAGWSDGWRVAIEVDGP
NCCL38 751 GPASPAATALASALAREAV-RRREGF---QIEELCQLWQVQQELGGKVAEALARSPDLNAAMEAAVAAER-ATKSTTSSTQEQVAEALKRLLQK------GCLPIGFVQTEVVVEGVLGRADIVAGWSDGRRVAIEVDGP
NCCL23 631 GPASPAASAVAPALAREAA-RRREGF---QTEALRQLWQAKHELGGEVTEALARSPDLLAAMEAAVAAER-ATGSTTSSLQEQVAEALQRLLRK------ARLPVVSVRTEVVVEGVLGRVDIVADWSDGRRVAIEVDGP
NCC1 660 GPASPAAAAMAPALAREAA-RRREGF---QTEGLLQLWQAQHELGGEVAEALARSPDLLAAIEAAVAAVQ-ATGSTTSSLQEQVAEALQRLLQK------GRLPIVSVRTEVVVEGVLGRVDIVADWSDGRRVAIEVDGP
NCCL24 666 VPASPAAAALAPALTREAA-RRREGF---PTGALSQLWQAQHELGGEVTEALARSPDLLAAMEAAVSAKR-ATGSTTSSLQDQVAEALQRLLQK------GRLPIVSVRMEVVVEGVLGRVDIVADWSDGRRVAIEVDGP
NCCL22 657 GPASPAAAALAPALAREAA-RRCEEL---QAEALSQLWQAQHELGGEVTEALARSPDLLAAMEAAVAAKR-ATGSNTSRLQEQVAEALQRLLQK------GRLPIVSVRTEVVVEGVLGRVDVVADWSDGRRVAIEVDGP
NCCL25 654 GPASPAAAALAPALAREAA-RRHEEF---QTEALSQLWQAQQELGGEVTAVLARSPDLLAAMEAAVAAER-VTGSNTSRLQEQVAEALQRLLQK------GRLPIVSVRTEVVVEGVLGRVDIVADWSDGRRVAIEVDGP
NCCL17 705 VPASPAAAALAPALAREAV-RRREEL---QTEALSQLWQAQHELGGEVAEALARSPDLLAAMEAAVADER-ATESNTSRLQEQVAETLQRLLQK------GRLPIVSVRTEVVVEGVLGRVDVVADWSDGRRVAIEVDGP
NCCL21 745 VPASPAAAALAPALAREAV-HRIRDL---AGDDFRQLWQAQHELGGEVAEALARSPDLLAAMEAAVAAQR-STTSTTSSTQKQVAEALQRLLQK------GRLPIVSVRTEVVVEGVLGRVDIVADWSDGRRVAIEVDGP
NCCL19 654 VPASPAASALAPALVREAV-RRREEF---QTEELRQLWQAQHELGGEVAEALARSPDLLAAMEAAVSAKR-ATGSTTSSTQEQVAEALQRLLQK------GRLPIVSVRMEVVVEGVLGLVDMVADWSDGRRVAIEVDGP
NCCL20 654 VPASPAASALAPALVREAV-RRREEF---QTEELRQLWQAQHELGGEVAEALARSPDLLAAMEAAVSAKR-ATGSTTSSTQEQVAEALQRLLQK------GRLPIVSVRMEVVVEGVLGRVDIVADWSDGRRVAIEVDGP
NCCL15 483 GPASPAAATLASALAREAA-RRREEF---QTEALSQLWQAQHALGGEVTEALARSPDLLAAMEAAVAAER-ATTSNTSSTQKQVAEALQRLLQK------DLLPIVSVQTEAAVEGVLGRVDIVADWSDGRRVAIEVDGP
NCCL35 410 VPASPAAAALAPALAREAV-RRREQFVDMSAEHLRQLWQAQHELGGEMAEALARSPDLNAAMEAAVAAER-ATGSNTSSTQKQVAKALQRLLQK------DLLPIVSLQTEPVVEGVLGRVDIVADWSDGRRVAIEVDGP
NCCL14 718 GPASPAAAALAPALALEAV-RRREEL---NGDGLRQLWQARQELGGEVAEALARSPDLLAAMEAAVAAER-ATTSTTSRLQEQVAEALQRLLQK------DLLPIVSVQTEVVVEGVLGRVDIVAGWRDGRSVAIEVDGP
NCCL1 658 GPASPAAAALAPALAYEAV-RRREKL---NGDDFSQLWQARQELGGEVAEALARSPDLLAAMEAAVAAER-ATTSTTSSTQEQVAEALQRLLQK------DLLPIVSVQTEVVVEGVLGRVDIVADWRDGRRVAIEVDGP
NCCL16
NCCL30 658 GPASPAASALAPALAREAV-RRRQEF---QTEGLLQLWQARQELGGEVAEALARSPDLLAAMEAAVAAER-ATGLNTSRLQEQVAEALQRLLQK------GRLPIVSVRTEVVVEGVLGRVDIVADWSDGRRVAIEVDGP
NCCL34 705 GPASPAASALAPALAREAV-RRREGF---QTEGLLQLWQAQQELGCEVAEALARSPDLLAAMEVAVVAER-ATGSTTSRLQKQVAEALQRLLQK------GRLPIVSVQTEVVVEGVLGRVDIVADWSDGRRVAIEVDGP
NCCL33 657 GPASPAAAALAPALAREAA-RRRQGF---QTEELRQLWQARQELGGEVAEALARSPDLLATMEAAVAAKR-ATESNTSRLQEQVAEALQRLLQK------GRLPIVLMQTEVVVEGVLGRVDIVADWSDGRRVAIEVDGP
NCCL28 667 GPASPAAATLASALAREAA-RRREEF---QTEALSQLWQARQELGGEVAEALARSPDLNAAMETAVAAEW-ATESNTSRLQEQVAEALQRLLQK------GRLPIVSVRTEVVVEGVLGRADMVADWSDGRRVAVEVDGP
NCCL29 663 VPASPAAAALALALAYEAV-RRREEV---QTEELFQLWQAQQELGGEVAEALARSPDLLAAMEAAVATQR-ATVSTTSSTQEQVAEALQRLLQK------GRLPIGSVQTEVVVEGVLGRVDIVADWSDGRRVAIEVDGP
NCCL27 647 GPASPAAAALAPALAREAV-RRREER---NGEYLCQLWQARQELGGEVAEALARSPDLLAAIEAAVAAKR-ATESTTSRTQEHVAEALRRLLQK------GRLPIVLVQTEVVAEGVLGRADIVADWSDGRRVAIEVDGP
NCCL31 653 GPATPAAAALAPALAREAV-RWREEL---NGDGLRQLWQARQELGGEVAEALARSPDLLAAIEAAVAAKR-ATESNTSRTQEQVAEALRRLLQK------GRLPIVLVQTEVVAEGVLGRVDIVADWSDGRRLAIEVDGP
NCCL32 617 GPATPAAAALAPALAREAV-RWREEL---NGDGLRQLWQARQQLGGEVAEALARSPDLLAAMEAAVAAER-ATGLNTSRTQEQVAEALRRLLQK------GRLPIVLVQTEVVAEGVLGRVDIVADWSDGRRLAIEVDGP
NCCL26 556 GPLSAGMQQLAVEVARDAA-GRWEEFT--ATEDLAQMWQAQQELGGEVAAALGSNGKLQAAMDAVVAARR-DNSKPLPDDHKQLLAALRRLEQHGGGATAGGLALESVQTGVVMSGVLAPLDAVVRLSDGRQVALELVGA
NCCL3 571 GTSSPGMQQLAIQLARNAA-GRWEGLI---HEDLSQLWQAQQELGGEVAETLCGISSLQAAMDKSVETYR-QDTKRLSETHKQLLAALRRLEQHQGREAAGGFAVVSVQAGVVAPGVLAAVDAVVRLNDGRQVAVELAGA
NCC2 568 ETGSPAMQQLAMQLARDAA-IRWEEFA---DEGLTQLWQAQQALGGEVAAALRGNRSLQAAMDKAVATYR-EDTKHLPDDQKQLLAALRRLEQHGRET-AGGLAVQSVQTGVVAQGVLTPVDAVMGLVDGRQVAVEMLGP
NCCL5 572 ETCSPDMRRLAIQLARDAA-SRWDDFT---IEELTQLWQAQQELGGEVAAALRDSSSLQATMDAVVAARR-EETKPLHANQKQLIGTLRRLEQHGRETAAGGLSIQSVQADAVVPGVLAPMAAVVGLSDGRQVAVEWIGL
NCCL6 435 GTSFPGMQQLAIELARDAA-IRWEGFT--AVAGLTQLWQAQRELGGEVAAALGGSPGLQAAMAAAVAAERAADVKPPPDNQKQLLAALRRLEQQGLITTTEGPSIQSLQTGVILPGILAPVAAVVGLSDGQQAAVELVGL
NCCL13 561 NTCSPGMQQLAVQLARNAAGRRWKGFT---DDGLRQLWQAQQELGGEVAAALGSSPGLQAAMDKSVETYR-QDTKRLSETHKQLLAALRRLE---GLATAGGLAVQSVQTGVVAPGVLTPVDVLVGLSDGRQVAVEWVTV
NCCL10 567 GTYSPDMQQLSIQLAREAV-SRWEGFT---TENLNQLWQAQQELGGEVAAALGSSPGLQAAMAAAVTAER-ENAKPTSDIQKQVVAALRRLEQHQGREAAGGLAVVSVQTGVVAQGALAPVDAVVRLSDGQQAAVELLGP
NCCL7 529 GTYSPDMKQLSIQLAREAV-SRWEGLT---TEHLKQLWQAQQELGGEVAAALGSGSGLQAAMAAAVTTER-EDAKPTSDTQKQVVAALRRLEQ-QGLKAAGGLAVQSLQTGVVAPGVLAPVDAVVRLSDGQQAAVEFLGP
NCCL8 426 GTYSPDMKQLSIQLAREAV-SRWEGLT---TEHLKQLWQAQQELGGEVAAALGSGSGLQAAMAAAVTTER-EDAKPTSDTQKQVVAALRRLEQ-QGLKAAGGLAVQSLQTGVVAPGVLAPVDAVVRLSDGQQAAVEFLGP
NCCL9 529 GTYSPDMKQLSIQLAREAV-SRWEGLT---TEHLKQLWQAQQELGGEVAAALGSGSGLQAAMAAAVTTER-EDAKPTSDTQKQVVAALRRLEQ-QGLKAAGGLAVQSLQTGVVAPGVLAPVDAVVRLSDGQQAAVEFLGP
NCCL11 484 GTSFPGMQQLAVQLARDAA-GRWEGFS---GEGHRQLWQAQQELGGEVAAALCSSPGLQAAMAAAVTAER-EDAKPTSDIQKQVVAALRRLEQHQGREAAGGLAVRSVQTGVFAPGVLAPVDAVVGLSDGRQVAVELVGP
NCCL12
consensus 841 gpaspaa la alarea rrre e l qlwqaqqelggevaealarspdl aameaavaa r at s ts q qvaeal rllqk g lpivsvqtevvvegvlgrvdiva wsdgrrvaievdgp


RAP Domain
NCCL2 537 IHFLRGRKG---NVSAVDGSTALRNRQLQRAFGKG-NVLYVPYWHWNGLKTPAKQEAYLLRRLQQ
NCCL36 833 DHFLTNRKD---DPSAVDGSTALRNRQLRRALGEG-RLLCVPYWEWYGLRTPSAQEAYLLQQLQDVLFGASSGAA-----AGEGIA--AAPRRQQQRTTQPQPDVAGSTTTTT--A------STSDTAGSSQKRRVL-FV
NCCL37 833 DHFLTNRKD---DPSAVGGSTALRNRQLRRALGAG-GLVCVPYWEWYGLRTPSAQEAYLLQQLQDVLFGASSGAA-----AGEGFP--AAPRRQQQRTTQPQPDVAGSTTTTTATA------STRNTAGSSQKRRVL-VV
NCCL38 880 DHFLANRKD---DPSAVGGSTALRNRQLRRALGEG-GLVCVPYWEWYGLRTPSAQEAYLLQQLQDVLFGASSGAA-----AGEGIA--AAPRRQQQRTTQPQADIAGSTTTST--V------STSDTTGSSQKRRVH-IV
NCCL23 760 DHFLTNRKD---DPSAVVGSTALRNRQLRRALGEG-GLLCVPYWEWNRRKTSSAQEAYLLQQLQDLLSGASSGAS-----AGEGSA--AARRRKEQRTTQPQPYTARSPTSASTTA---------IAAGSSHKRRIL-VV
NCC1 789 DHFLTNRKD---DPSAVNGSTALRNRQLQRALGEG-GLVCVPFWEWYGLRTPSAQEAYLLQRLQDLLSGASSGAA----G-GGGSA--AAPRHQQQRTTRLQPDIAGSSTTS-----------TTSTANSSQKRRVL-VV
NCCL24 795 DHFLTHRKD---DPSAVIGSTALRNRQLRRAFGEG-GLLCVPYWEWDRQKTSSTQEAYLLQQLQDLLSGASSGTA----AAGEGSA--AARRRQQQRTTRLQPDTANSSTTTSTTT------ATTTTAGSSQKRRIL-VV
NCCL22 786 DHFLIDRKD---DPSAVVGSTALRNRQLRRALGEG-GLLCVPYWEWDRQKTSSAQEAYLLQRLQDLLSGASSGAS-----AGYGTA--AARRRKEQRTTRLQPDTANSPTSASTAA------STSTTASSSQKRRIL-VV
NCCL25 783 DHFLTNRKD---DPSAVDGSTALRNRQLRRAFGEG-GLLCVPYWEWDRQKTSSAQEAYLLQQLQDLLSGASSGAS-----AGEGSA--AAPRHQQQRTAQPQPDTANSSTT-----------STTSTASSSHKRRVL-IV
NCCL17 834 DHFLTNRKD---DPSAVDGSTALRNRQLRRALGEG-GLLCVPYWEWYERQAPSAQKAYLLQQLQDLLSGASSGAS-----AGEGSA--AARRRKEQRTTQPQPYTAGSNTSASTTASIT------TTASSSQKRRIL-VV
NCCL21 874 DHFLTNRKD---DPSAVDGSTALRNRQLRRALGEG-GLLCVPYWEWYERQAPSAQEAYLLQQLQDLLSGASSGAS-----AGEGSA--AARRRKEQRTTQPQPYTARSPTSASTTASITTTATTTTTASSSQKRRIL-VV
NCCL19 783 DHFLTHRKD---DPSAVIGSTALRNRQLRRAFGEG-GLLCVPYWEWYGLRTPSTQEAYLLQQLQDLLSGASSGTA-----AGKGSA--AAPRHQQQRTAQPQPDTANSS----------TTATTTTTASSSQKRRIL-VV
NCCL20 783 DHFLTHRKD---DPSAVIGSTALRNRQLRRAFGEG-GLLCVPYWEWYGLRTPSTQEAYLLQQLQDLLSGASS*
NCCL15 612 DHFLTNRKD---DPSAVDGSTVLRNRQLRWAFGEG-GLLCVPYWEWDRRKTYSAQQAYLLQQLQDLLSGASSGAA----SAGEGSA--ATPRRKQQRTTPPQPDTAGRSSSTTNTT----STSASTTAGSSQKRRVLVVV
NCCL35 542 DHFLTTRKD---DPYAVIGSTALRNRQLRRAFGRG-GLLCVPYWEWDRRKTPSAQEAYMLQQLQDLLSGASSGA-----AAGEGSA--ATPRRKQQRTIPPQPDTAGSSSITTATT--------STTASSSQKRRVLV-V
NCCL14 847 DHFLTNRKN---DPSAAIGSTALRNRQLRRAFGEG-GLLCVPYWEWYGLKTSSAQEAYLLQQLQDLLSGASSGAV----AAGEGSA--ATPRRKQERTTRPQPNTAGRTSTS-----------T-TAGSSQNRRVLVVV
NCCL1 787 DHFLTNRKD---DPSAVNGSTVLRNRQLRRAFGQG-GLLCVPYWEWDRRKTSSAQEAYMLQQLQDLLSGASSGA-----AAGEGSA--AAPRRPQQRTTPPQPDTAGSSTTT-----------STTASSSQKRRVLVVV
NCCL16
NCCL30 787 DHFLTNRKD---DPSAVDGSTALRNRQLRRAFGEG-GLLCVPYWDWCGRKTPSFQEAYLLQQLQDLLSGAPSGAA----AGGEGSA--AAPRRKQQRTTQPQPDTAGTTTT-----------ATTTASSSQKRRIL-VV
NCCL34 834 AHFPTNRKD---DPSAVIGSTALRNRQLRRAFGEG-GLVCVPYWEWYGLRTPTAQEAYLLQRLQDLLSGASSGAA----AGGEGIA--AAPHCKQQRTTRPQPDTAGSSTTTASTT-----TITSTTAASSEKRRIL-VV
NCCL33 786 DHFLTNRKD---NPSAVDGSTALRNRQLRRTLGEG-GLVCAPYWEWYSLRTPSAQEAYLLRRLQDLLSGASSGGA----AAGEGIA--AAPRRQQQRTTRPQPDTAGTTTS-----------TSTTAVSSQERRIL-VV
NCCL28 796 DHFLTNRKD---DPSAVDGSTALRNRQLRRAFGEG-RLLYVPYWEWDIRKTPTAKEAYLLQRLQDLLSGASSDTA-----AGKGNA--AAPRRQQQRTTQPQPDTAGTTTT-----------ASTTASSSQKRRIL-VV
NCCL29 792 DHFLTNRKD---NPSAVDGSTALRNRQLRRTLGER-GLVCAPYWEWYSLRTPSAQEAYLLRRLQDLLSGAPSGAA----AGGEGSA--AAPRRKQQRTTRPQPDTAGTTTT-----------ATTTASSSQKRRIL-VV
NCCL27 776 DHFLTNRKD---DPSAVVGSTALRNRQLRRAFGEG-GLLCVPYWEWYGRKTSSAQEAYLLQQLQDLLSGASSGAA----A-GEGSA--AARRRKQQRTTRPQLDTAGITTST---T-----TSTTTTAGSSQKRRIL-VV
NCCL31 782 DHFLTNRKD---DPSAVIGSTALRNRQLRRAFGEG-GLLCVPYWEWDDRQTPSFQEAYLLQRLQDLLSGAPSGAA----AGGEGSA--AAPRRKQQRTTRLQPDTAGTSTA-----------ASTTASSSQTRRIL-VV
NCCL32 746 DHFLTNRKD---DPSAVIGSTALRNRQLRRAFGEG-GLVCVPYWEWDERQTPSFQEAYLLQRLQDLLSGAPSGAA----AGGEGSA--AAPRRKQQRTTQPQPDTAGTTTT-----------ATTTAVSSQTRRIL-VV
NCCL26 692 VRFLSNRR---RDPAAVNGSTAMRHRHVRREYGEG-GVLMVPYWEWESLTSPEEQEAYLLRRLQALLLATAQRAV-------------GTPTRTTAG--D--AASPT-----------------SKSSSCGSPQLL--A
NCCL3 706 VSFLANLRQ--RDTAAVNGGAAMRHRQLRRAFGEG-GVLLVPYWEWDRLQTAEEQEAYLLRRLQKVVVMAAEPQAAEAATAGAGRAAAGAPVRTTGA--TASVVSPSSNNSNSS--------RGSPGGGSPGRSRPL--A
NCC2 702 KRFIYNRKQ--DDPTAVNGGTGMRNRQLRRAFSEG-GVLLVPHWEWEGLKSPEEQEAYLLLRLQEVAAA-AETQAAAAAAQGAGKTSARTSAAGAAA--APAPARPGSSGSSSS--------RGGRGGGSPGSPQLL--T
NCCL5 707 KRVLRSRQR--EGAILLEGGIGMRIRQLQRAFREG-GVLVVPHWEWEGRRSPEEQETYLLRRLQEVVAA-AETQAAAAAAKGAGKAPAGTPAAGAAA--APVPARPGSSSSS-------------PGSGSPGSPQLL--A
NCCL6 572 KRYLYNRKQPQ-DPADVDGGTAMRIRQLRRAFGEGGGVLVVPYWEWAGLKSPEEQEAYLLHRLQEPMAPPPEVVG-------NGGI--RASE--ETR--SAAP---AASSPAAA--------TSGDGAAGNK*
NCCL13 694 VRFLSNRKQ---DPSAVNGSTVLRHRQLRRAFGEG-GVLLVPYWEWDGLQTAEEQEAYLLRRLQQPAVAVETVT--------AGVV--GGAA--AVI--TAAPHQPATTTTTTT--------KGGGGSGSNPQQQLL--V
NCCL10 702 ARFLSNLKR---DPTAVDGGTAMRIRQLRRAFGEGGGVLVLPYWEWAGLKSPEEQEAYLLRRLQQPAVAVETVA--------VGAV--GGAA--AVI--TAAPQQPATTTT-TS--------KGGGGSGGNPQQQLL--V
NCCL7 663 KRYLLTQTQQQDPSAAVEGGTAIRIRQLRRAFKEG-GVLLVPYWEWDFL-TPSEQEAYLLRRLQQPAVAVETVT--------AGAV--GGAA--AVT--TAAPHQPATTTTNSN--------KGGGGSGGNPQQQLL--V
NCCL8 560 KRYLLTQTQQQDPSAAVEGGTAIRIRQLRRAFKEG-GVLLVPYWEWDFL-TPSEQEAYLLRRLQQPAVAVETVT--------AGAV--GGAA--AVT--TAAPHQPATTTTNSN--------KGGGGSGGNPQQQLL--A
NCCL9 663 KRYLLTQTQQQDPSAAVEGGTAIRIRQLRRAFKEG-GVLLVPYWEWDFL-TPSEQEAYLLRRLQQPAVAVETVT--------AGAV--GGAA--AVT--TAAPHQPATTTTNSN--------KGGGGSGGNPQQQLL--A
NCCL11 619 TRFLSDQTCS--GRSAVDGRTVLRIRQLRRAFGEGGGVLVVPYWEWDALTN-AEEKAYLLRRLQQPAVARAVTV---------GAE--GGAA--AVT---------AAATTTTA--------KGGGGSGGNPQQQLL--V
NCCL12
consensus 981 dhfltnrkd dpsav gstalrnrqlrrafgeg gllcvpywew ktps qeayll lqdvl ga s g g a aa rr rt qp agsttt a ss rrvl v

```
NCCL2
NCCL36     953 RRKAPKQEEAAGSGG-AGGASGATATA-----AAGAELVAPAGQRQAVEAQRLPNSGAGGGAGEGGGDGTPLAPQPPPPQ-RQR--------VSAPRARRNSSQSRSRSSQAGSPLPPP-PSPVKAPPPEQVQ--AAAPP
NCCL37     955 RRKTPKQEA-PGGGG-AGEASGATAAA-----AAGAELAAPAGQRQAEEAQRLPSLGASGGAGEGGGDGTPLAPQPPPPQ-RQR--------VSAPRARYSSSQSRSRSSQAGSPSPAP-P-PVEAPPPEQVQ--AAAPP
NCCL38    1000 RRKTPKQEA-AGGGG-AGGASGATAAA-----AAGAELAAPAGQRQAEEAQRLPSSGAGGGAGEGGGDGTPLAPQPPPPQ-RQR--------VSAPRARYSSSQSRSRSSQAGSPSPAP-P--VKAPPPEQVQAQAAAPP
NCCL23     879 RGKAPK-QEAVDGGG-AGGASNATAAAA-TVAVAGAEVVAPAGQQQAEESHQLPNIGAGGGAGEGGEDGTPLAPQPPPPQ-QQH--------VSAPRARRSSSQSRSRSSQAGSPSAPP-PPPVEAPPPEQMQ--AAAPP
NCC1       906 RRKALK-QEAAGGGG-AGGASNATAAAA-TVAVAGAESVAPAWQQPVEKAQRLPSSGAGGGAGEGGEDGASLAPQPPPPQ-QQH--------VSAPRARRSSSQTQSRSRSS*
NCCL24     918 RRKAPKQQEAAGGGG-AGGGSDATAAAV-AAGVAGAESVAPAWQQPVEKAQRLPSSGAGGGAGEGGEDGASLAPQPPTPQ-QQR--------VSAPRARRSSSQSQSRSSQAGLPAPAPAPPSVKAPPLEQTR--QLNPG
NCCL22     908 RKKAPM-QEAAGGGG-AGGASGATATAA-TI-PAEAKVVAPALQQQAEESHQLPNIGAGGGAGEGGGDGTPLAPQPPPPQ-QQR--------VSAPRARRSSSQSRSRSSQAGSPSPPPPPPPVQAPPPEQMQ--VAAPP
NCCL25     900 RRTARK-QEAAGGGG-AGEASGATAAAR-TV-RAEAEVAAPARQQSAEKAQRLPSSGAGGGAGEGGGNGTPLAPQPPPTQ-QQR--------VSAPRARRSSSQSRSRSSQSGSP--PP-APPVEAPPPEQTQ--AAAQS
NCCL17     956 RRKAPK-QEAVGGGG-AGGASSATAAAA-TI-PAEAKVVAPALQQQAEESHQLPSSGAAGGAGEGGGDGTPLALQPPPPQ-QQH--------VSAPRARRSSSQSRSRSSQAGSPSAPP-PPPFEAPPPEQVQ--AAAPP
NCCL21    1002 RRKAPK-QEAVGGGG-AGGASSATAAAA-TI-PAEAKVVAPALQQQAEESHQLPSSGASGGAGEGGGDGASLALQPPPPQ-QQH--------VSAPRARRSSSQSRSRSSQAGSPSAPP-PPPFEAPPPEQVQ--AAAPP
NCCL19     901 RRKAPK-QEAAGGGG-AGEASGATAAAR-TV-PAEAEVAAPARQQSAEKAQRLPSSGAGGGAGEGGGNGTPLAPQPPPPQ-QQR--------VSAPRARRSSSQSRSRSSQAGSPS--P-APPVEAPPPEQVQ--AAAPP
NCCL20
NCCL15     738 RRKAPKQEAVGGGGG-AGGASDATVDAA--TVPAEAEVAAPARQQQAEESQRLPISGAGGGAGEDGGDGAPLAPQPPPPQQRPR--------VSAPRARRSSSQSHSSSSSSQA--GSPPPPSVEAPPPEQVQ--AAAPP
NCCL35     662 RRTAPKQEAVGTSGG-AGGASDAAAAATATAAAAGAKSVAPAWQQQAEEAQRLPSS---GGAGEGGGDGAPLAPQPPPPQ-QPR--------VSAPPTRSSSRQS*
NCCL14     964 RRKARKQQEVACGGG-AGEASDATVAAA--TVPAEAEVVAPAGQQQTETVQRLPNSGAGGGAGEGGGDRAPLAPQPPPPQ-QPR--------VSAPRARRSSRQSPSRS---IQ--AGSPPPSVEAPPPEQMQ--AAAPP
NCCL1      904 RRTA-PKQEAAGGGG-AGEASDATVAAR--TVPAEAEVVAPAGQQQTETVQRLPSSGAGGGAGEGGGDGAPLAPQPPPPQ-QPR--------VSAPRTRRSSRQSQSRG---VQ--AGLPAPSVEAPPPEQMQ--AAAQP
NCCL16
NCCL30     904 RRKAPKQDLQHDATCDAPSASGAGPAE------AVAEVA---------------TSG---GP-RPATAGADVASAPRRQQRQRQL--QQQPTRPEPEAVAAAV-GRRRRPAAK---EPRPQQQQTRPPPQNTRKRAASAN
NCCL34     958 RRKARKQDLQHNATFDTPSASGAGPAE------AVAEVA---------------TSG---GL-RPAAAGADVAPAPRQQQRQRQ---HQQPTRPEPEAVAATV-SRRRRPAAN---EPRPQQ-QQRPPPQNTRKRAAASN
NCCL33     903 RRTARKQDLQHDATCDTPSASGAGPAE------AVPEVA---------------TSG---GL-RPAAAGADVAPAPRQQQRQRQQ*
NCCL28     912 RRKAPKQDLQHDATCDAPSASGAGPAE------AVAEVA---------------TSG---GL-RPAAAGADVASAPRQQQRQRQ---LQQPSRPEPEAVAVAV-GSCSRPAAN---EPRPQQ-QTRPPPRNTRKRAASAN
NCCL29     909 RRTAPKQDLQHDATCDAPSASGAGPAE------AVAEVA---------------TSG---GL-RPAAAGADVASAPGRQQRQRQ--RQQQPSMPEPEAMAAAV-GRRRRPAAN---EPRPQQQQTRPPPQNTRKRAASAN
NCCL27     896 RRTARKQDLQHDATCDAPSASGAGPAE------AVAEVA---------------TSG---GP-RPATAGADVAPAPRQQQR------QQQPSRPEPEAVAAAV-GRRRRPAAK---EPRPQQQQERPPPRNTRKRAAFAN
NCCL31     899 RRKAPKQDLQHDATFDAPSASGAGPAE------AVAEVT---------------ASG---GS-RPAAAGADVASAPRRQQRQRQLQQQQQPSRPEPEAVAAAV-GRRRRPAAK---EPRPQQQQERPPPQNTRKRAASAN
NCCL32     863 RRKAPKQDLQHDATCDAPSASGAGPAE------AVAEVA---------------TSG---GP-RPAAAGADVAWAPGQQQQQRQ-RQLQQPSRPEPEAVAAAV-GRRRRPAAK---DPQPQQQQTRPPPQNTRKRAAASN
NCCL26     791 R-PPRR-VP*
NCCL3      831 R-SPRR-V*
NCC2       826 R-PPRR-RV*
NCCL5      826 R-PPRR-RV*
NCCL6
NCCL13     806 R-PARR-PEGGGGG*
NCCL10     814 R-PARR-PEGGGGGGGSGRGGR*
NCCL7      777 R-PARR-PEGGGGRRRVGGS--KRTVGLRS*
NCCL8      674 R-PARR-PEGGDGGGGGGGGSKRTVVGLRS*
NCCL9      777 R-PARR-PEGGDGGGGGGGGSKRTVVGLRS*
NCCL11     724 R-PARR-PEGGGGGGGGSGRGGR*
NCCL12
consensus 1121 r  a k     g   a  as a         a                     g     g   la  p   q
```

```
NCCL2
NCCL36    1075 RRAAGEKRTERKAAAAPAAGGGG--GGGGGSDPGAVVVPAAVSEKQVAVVAGEEGSTAAAVKGGD--GGGSGGDGGLPAGSRRRRAVGTGRRAASSSLAAATEQMRLEQ*
NCCL37    1075 RRAAGGKRTERKAAAAPAADGGG--GGGGGTDPGAVVVPAAVSEKQVAVVAEEEGSTAAAVKGGD--GGGSGGGGGLPAGSRRRRVVGTGRRAASSSLAAASEQMRLEQ*
NCCL38    1121 RRAAGGKQTV---AVDPAAGGGG--GGGGGTDPGAVVVPAAVSEKQVAVVAGEEDSSAAAVRGGD--GGGSGGGGGLPAGSRRRRAVGTGRRAASSSLAAASEQMRLEQ*
NCCL23    1004 RRAAGGKRTERKAAAATAAGGGG--GGGGGSDPGAVVVPAAVSEKRLSVVAGEADSSAAAVECGGGH----GSGGGLLAGSRRKRAVGTGRRAAQKPSSSE*
NCC1
NCCL24    1045 GPRVGSGLSGKQ---------------RRPRQQAAVAAAAAVTPGRWWFRPRSLRSDS----------------RWWPGRRTRRQQL*
NCCL22    1033 RRAAGGKWTERKAAAAPAAGGGG--GGGGGSDPGAVLVPAAVSEKR*
NCCL25    1022 RRAAGGKRTERKAGVAPAAGDG----GGGGSDPGAVVVPAAVSEKQVVVVAGEADSLAAAVECGGG-GDGSGNGGGLLAGSRRGQAVGTGRRAAQKPSSSE*
NCCL17    1080 RRAAGGKRTERKAAAAPAAGGGGGGGGGGSDSGAVVVPAAVSEKRLSVVAGEADSSAAAVECGGGGDGSGSGGGLPAGSRRGQAVGTGRRAAQKPCSIE*
NCCL21    1126 RRAAGGKRTERKAAAAPAAGGGG-GGGGGGSDSGAVVVPAAVSEKR
NCCL19    1023 RRAAGGKRTERKAAAAPAAGGGG---GGGGSDPGAVVVPAAVSEKRLSVVAGEADSSAAAVECGGGG---DGSGGGLPAGSRRGQAVGTGRRAAQKPCSIE*
NCCL20
NCCL15     863 RRAAGGKRTERKAAAAPAAGGGG----GGGSDPGAVVVPAAVSENQVAVVAGEEGSTVAAVAGGG-GGGGSGGGGGLPAGSRRRRAGGTGRRAAQKSSSSV*
NCCL35
NCCL14    1085 RRTAGGKRTVRKAAAAPAAGG-------GGSDPEAVVVPAAVSEKQAAVVAGEEGSTLADVEGGG-SGSGRGGGSRRR-----RAV------GPAGVRPRQA*
NCCL1     1024 RRAAGGKRTERKAAAAPAAGGGG--GGGGGSGPGAVVVPATVSENQVAVVAGEEGSTAADVEGGG-GGGGSGGAGACP-----RAAGGGGRRGPAGVQPRSVPGTSE*
NCCL16
NCCL30    1013 A-------HAAAAAAAAAA*
NCCL34    1065 A-------DAAAAAA*
NCCL33
NCCL28    1019 A-------HAAAAAAAAAA*
NCCL29    1018 A-------HAAAAAAAAAA*
NCCL27    1001 A-------HAAAAAAAAAA*
NCCL31    1010 A-------HAAAAAA*
NCCL32     973 A-------HAAAAAA*
NCCL26
NCCL3
NCC2
NCCL5
NCCL6
NCCL13
NCCL10
NCCL7
NCCL8
NCCL9
NCCL11
NCCL12
```

B)

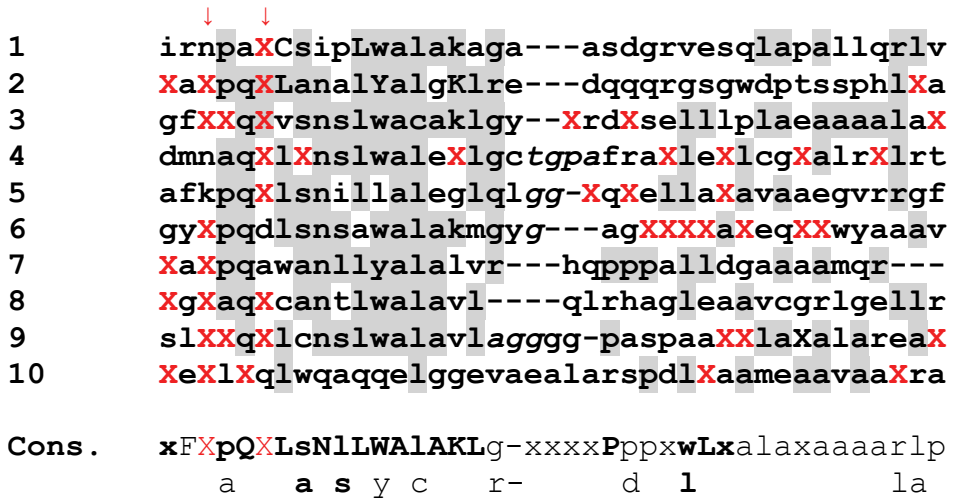


Figure S2: Alignment of NCCL proteins.

A) Alignment was done with the clustalΩ 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. 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.

Intracellular targeting of NCL proteins.

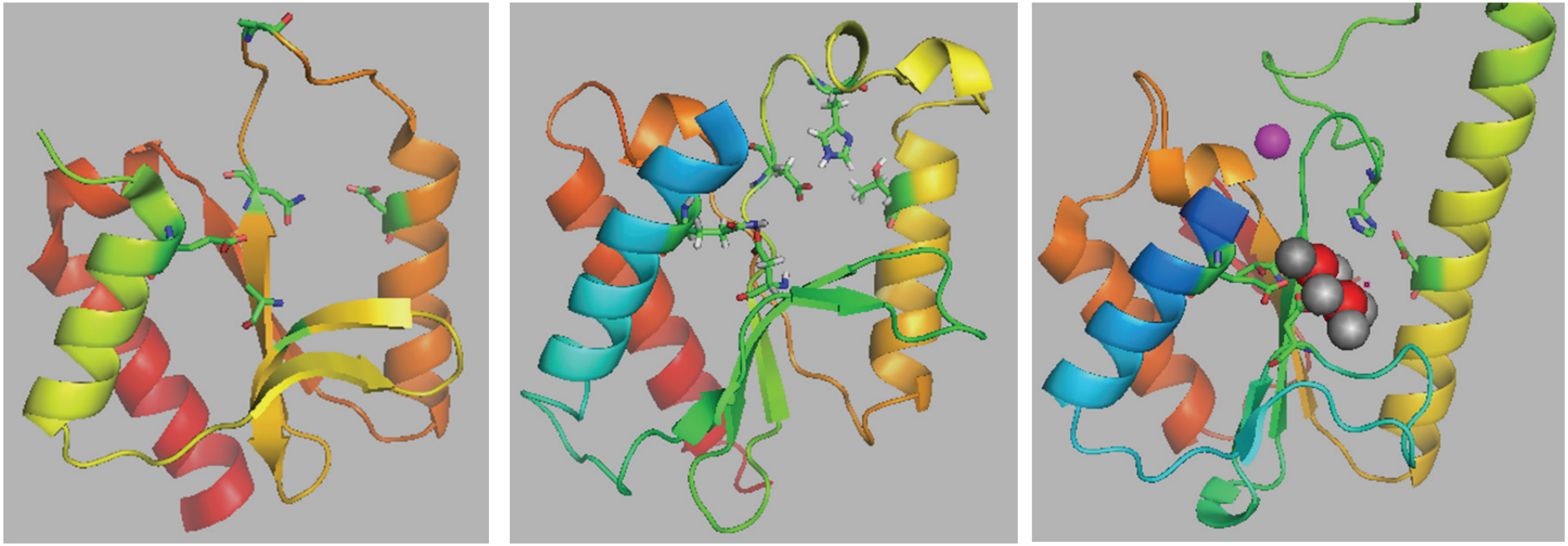
All NCL proteins are predicted to be targeted to an organelle by at least 3 of the four programs (Predotar: https://urgi.versailles.inra.fr/predotar/predotar.html (Small et al., 2004); Predalgo: https://giavap-genomes.ibpc.fr/predalgo (Tardif et al., 2012); TargetP: http://www.cbs.dtu.dk/services/TargetP/ (Emanuelsson et al., 2001) and WoLF PSORT II: http://www.genscript.com/psort/wolf_psort.html (Horton et al., 2007)). WoLF PSORT II predicts mostly a chloroplast localisation for NCL proteins, but the other programs predict mostly a mitochondrial targeting. Whereas this may reflect a possible mitochondrial localisation for some NCL proteins, this may also result from the poor performances of these programs towards *Chlamydomonas* proteins. Indeed, known chloroplast localised proteins, such as, for example the CCB1-4 factors, which bind a c-type heme to the chloroplast-encoded apocytochrome b_6 (Kuras et al., 2007), were also predicted as targeted to the mitochondria.

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 αββββαβ 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.

Fig. S4: Comparison of NCC1 model with known structures of endonucleases

Structure of 3r3p **Model generated for NCC1 by I-TASSER** **Structure of 1cw0 (protein + metals)**

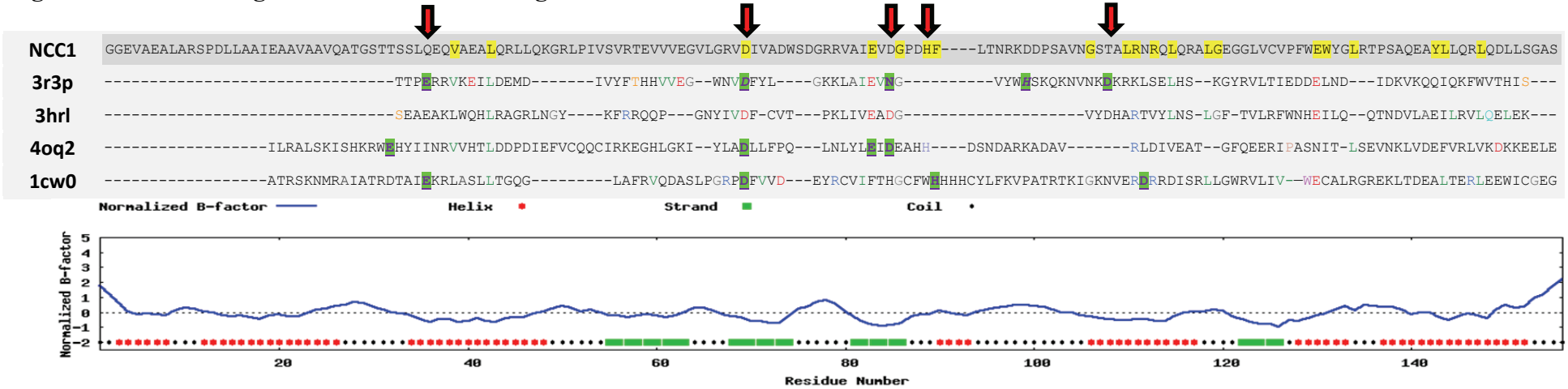


(res 1-31 removed)

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. S5: I-TASSER alignment used for threading of the NCC1 RAP domain.



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 SI: Genetic independence of the *ncc2* mutation from *MCA1* and *TCA1* genes.

	phenotype	Number of descendants
A	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

Table EII: Markers designed to map the ncc2 mutation on chromosome 15

marker	primers	position on chromosome 15	Ann. T. (°C)	elong. (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 primers, 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 our *C. reinhardtii* laboratory strains derived from *137c* and in S1D2. - indicates that no PCR product could be amplified from S1D2 with the indicated pair of primers.

[a]: oligonucleotide with 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 S1D2.

Supplemental Table SIII: 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$_{Fus}$FW	CGC**AAGCTT**CAATGCGTACTCCAGAAGAACT	***HindIII***
atpA$_{Fus}$RV1	CGC***CTGCAG***ACCATCACCTAATAATACCGCA	***PstI***
atpA$_{Fus}$RV2	GCT***CTGCAG***TTAAGCTGTCATACTACCTTCA	***PstI***
*atpA*Cod	CAATGCGTACTCCAGAAGAACT	
*ncc1*_FW	CGTTCATTCCGAGACACCTG	
*ncc1*_RV	TGCGGCTACGGCTTTGGGTT	
*ncc2*_FW	TATGACACCACAGGCACTCA	
*ncc2*_RV	CCAGACCTTGCTCAGTCAGT	
*atpA*T1^{M}FW	AACTCAGGTTTACGCCCAGCG**GTTAAC**GTAGGTATTTCAG	***HpaI***
*atpA*T1^{M}RV	CTGAAATACCTAC**GTTAAC**CGCTGGGCGTAAACCTGAGTT	***HpaI***
atpA$_{Ext}$FW	AGGTAGTATGACAGCTCTTC**CAATTG**TTGAAACACAAGAAGGT	***MfeI***
atpA$_{Ext}$RV	TATACCGAGTTCCTAC**TTAATTAA**AAATTAAGCAGCTTTAGCT	***PacI***
*petA*T2^{M}FW	ACAAAACTACGCTAACC~~CCCGA~~GAAGCAAACGGTCGTATTGTATGTGCA	~~*PmlI*~~
*petA*T2^{M}RV	TGCACATACAATACGACCGTTTGCTTC~~TCGCG~~GGTTAGCGTAGTTTTGT	~~*PmlI*~~
petA$_{Ext}$FW	TCCCCATTTTTATAA**AGATCT**TCCATGCATGAACT	***BglII***
petA$_{Ext}$RV	TGGGATTTT**GTCTAC**AACAACTTCACCGTTTGCTTTT	***AccI***
*petD*5::T2_FW	TAG**ATTTAAAT**CCACGTGAGGCTAATGGTGTAATTCTGTCCCTTTTTAC	***SwaI***, *PmlI*
petD$_{Cod}$::T2_RV	GGC**CTGCAG**CCATTAGCCTCACGTGGAATACACCTAGAAGTTTGTTTG	***PstI***, *PmlI*

[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 SIV: *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	Cre15.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 gene model written in green is that of the OPR protein most closely related to these paralogs that was used as an out-group to root the phylogenic tree in Fig. 5D.

Table EV: the NCL gene family.

Name[a]	Name[b]	Genomic_position[c]	EST-supported gene model[d]	Comments[e]	Nb of introns[e]	exons junction[f]
OPR_68	Cre15.g634800.t1.1	15:92140-97565_rev.			2	RAAFHLTGK LSSGSRGVQ SLWALAVLG SVFNQQELS
NCL1	Cre04.g223150.t1.1	4:2669273-2673984_for.	ova_1_g4772.t1		1	SAAFNLCGK LESARTSGP
NCL2	Cre06.g258051.t1.1	6:1256013-1258462_rev.	ova_1_au5.g12558_t1	inactivated gene; not expressed at the RNA level	0	
NCL3	Cre15.g640350.t1.2	15:686690-690305_rev.			1	WRAFNKAAK LRSGDGGAA
NCC2	Cre15.g640400.t1.1	15:690920-695055_rev.			1	SNAFRKAVQ YRSGNGGAA
NCL5	Cre15.g640450.t1.2	15:697680-700845_rev.			1	SNAFRKAVQ LRSADGGAA
NCL6	Cre15.g639850.t1.2	15:804529-808224_rev.	ova_1_au5.g5854_t1		1	FAAFNLCGR LKSRAGAAA
NCL7	Cre15.g639o800.t1.1	15:808469-812224_rev.	Chlre_OPR7		1	CAAFNLCGK LESRAGASQ
NCL8	Cre15.g639802.t1.1	15:812279-815974_rev.	Chlre_OPR8	Missing N-ter (sequence gap)		CAAFNLCGK LESRAGASQ
NCL9	Cre15.g639750.t1.1	15:816029-820024_rev.	ova_1_g13717.t1		1	CAAFNLCGK LESRAGASQ
NCL10	Cre15.g639700.t1.1	15:820079-824234_rev.	ova_1_g13716.t1		1	CAAFNLCGK LESRAGASQ
NCL11	Cre15.g639650.t1.1	15:824443-827924_rev.			1	SAAFNLCGR LESSGGPAA
NCL12	Cre15.g639614.t1.2	15:827979-830794_rev.		Truncated NCL protein	1	CAAFSKCGN LESRSAGAA
NCL13	Cre15.g637850.t1.1	15:878963-882816_for.			1	CAAFNKCGK LESRAGANQ
NCL14	Cre15.g639308.t1.2	15:927920-932861_for.	ova_1_g13706.t1		1	SAAFNLCGK LESARGGGP
NCL15	Cre15.g639304.t1.2	15:932910-937551_for.	ova_1_g13705.t1		1	SAAFGRCGK --SARAGGP
NCL16	Cre15.g639300.t1.1	15:938660-941367_for.	Chlre_NCL16	Truncated NCL protein; maybe a pseudogene	2	SLAPPGPPG SGGGLRLL SAAFNLCGK LDSARAGGP
NCL17	Cre15.g639000.t1.2	15:992161-997409_rev.			1	SAAFGRCGK LESARAGGP
NCC1	Cre15.g638950.t1.1	15:997456-1003259_rev.			1	SAAFGRCGK LESARAGGP
NCL19	Cre15.g638954.t2.1	15:1003810-1010065_rev.			3	SAAFGRCGK LEPARAGG
NCL20	Cre15.g638956.t1.1	15:1011429-1017745_rev.	ova_1_g13694.t1			SAAFGRCGK LESAHASGP
NCL21	Cre15.g638750.t1.2	15:1024446-1029775_rev.	Chlre_OPR21	Inactivated gene	1	SAAFGRCGK LESARAGGP
NCL22	Cre15.g638700.t1.1	15:1029830-1034285_rev.	ova_g13692.t1		1	SAAFGRCGK LESARAGG
NCL23	Cre15.g638650.t1.1	15:1034333-1039095_rev.			1	SAAFNLCGK LESARAGGP
NCL24	Cre15.g638651.t1.1	15:1039144-1044255_rev.	ova_1_g13690.t1		1	SAAFGRCGK LESARAGGP
NCL25	Cre15.g638550.t1.1	15:1044303-1048895_rev.	ova_1_g13689.t1		1	SAAFGRCGK LEPARAGGP
NCL26	Cre15.g638450.t1.2	15:1063996-1067484_for.	ova_1_au5.g5827_t1		1	SNAFRKAVQ LRSGDGGAA
NCL27	Cre15.g638401.t1.1	15:1071566-1076437_for.			1	SAAFNLCGK LESARAGGP
NCL28	Cre15.g638300.t1.1	15:1076486-1081385_for.			1	SAAFNLCGK LESARAGGP
NCL29	Cre15.g638303.t1.1	15:1081876-1086727_for.	ova_1_au5.g5823_t1		1	SAAFNLCGK LESARAGGP
NCL30	Cre15.g638304.t1.1	15:1086776-1092814_for.	ova_OPR74		1	SAALGRCGK LESARAGPA
NCL31	Cre15.g638150.t1.1	15:1093526-1097989_for.			1	SAAFNLCGK LESSHAGPA
NCL32	Cre15.g638100.t1.1	15:1098486-1102786_for.	ova_1_au5.g5820_t1		1	SAAFNLCGK LESSHAGPA
NCL33	Cre15.g638050.t1.1	15:1102836-1107831_for.			1	SAAFNLCGK LESAHAGGP
NCL34	Cre15.g638000.t1.1	15:1108216-1113927_for.			1	SAAFNLCGK LESSHAGGP
NCL35	Cre17.g739800.t1.1	17:5864782-5869727_rev.	Chlre_NCL35		2	SAAFNLCGK LESARAGGP GAALRQLRT PKEAEAFKP
NCL36	Cre19.g750497.t1.1	s19:62092-66415_rev.			1	SAAFGRCGK LESACAGGP
NCL37	Cre19.g750697.t1.1	s19:102916-108047_for.	ova_1_g16839.t1		1	SVAFNLCGK LESACAGGP
NCL38	Cre19.g750747.t1.1	s19:108096-113305_for.	ova_1_g16838.t1		1	SVAFNLCGK LESACAGGP

[a]: NCL genes were numbered based on their position on the genome. Same colour code as in Table EIV. The more divergent gene model ova_1_g13706.t1, located on chromosome 15 but outside of the cluster, was called Chlre_OPR68. The colour of the thick line on the right remind of the subclusters identified in Fig 5 A, D

[b]: Phytozome v5.5 gene models available from http://phytozome.jgi.doe.gov/

[c]: Position of v5.5 gene models on the genome (format "chromosome number":"start"-" end"_"orientation"). s19 refers to the scaffold 19.

[d]: Some gene models were improved to fit the transcriptomic data available at http://genomes.mcdb.ucla.edu/Cre454/. ova_1_xx gene models can be found in the "user models" track on the Chlamydomonas genome v4.0 brower: http://genome.jgi-psf.org/Chlre4/Chlre4.home.html. Gene models Chlre_NCCLxx gene models are described below.

[e]: Additional comments regarding some NCL proteins.

[f]: number of intron within *NCL* coding sequence.

[g]: Sequence of exon junctions, illustrating the conserved position of the intron (see also the alignment in Fig E3).

```
>Chlre_NCL7
     mRNA            complement(join(<812060..811548,811328..>809433))
                     /gene="Chlre_NCL7"
                     /product="predicted NCL protein"
     CDS             complement(join(812060..811548,811328..809433))
                     /gene="Chlre_NCL7"
                     /codon_start=1
                     /product="predicted NCL protein"
MLALRVVAGSPAGAHRHSVHGCSQSLSARLQAPPPACAFRRLFVYGMHRTAADTGPSSPRAAAATSVPAL
NSSSSSPLRVSTAVVLCGSSSSRLGRPIVQAAAGRGSGGGRAGDGSRASGRSRGSGRDPADDMTVCRTLE
ELQAIISQRLSVWEGRKDFITMCAAFNLCGKLESRAGASQTAVARSSIIGMLAPAYLPLVPRIRNAKNCS
IPLWALGKAGLVGGGTEAQLAAALLERLVDPEVLGKAEPQNLAIVMYALGKLKQDQQQRGSGWDPTSSPH
LQTLAAAVASRLRTVAGHGFKPQDVSNSLWALKALGCTGPEYRSAVEALCGEALQQLRKPELAAAFIPQQ
LSNILLALEGLQLCSEQAELVVAVAEEDMRRGFDGYIAQDLSNSAWALAKMGFGAGPEAPAEQRQWVSAA
VDAATQSGAMARATSQAWSNLLYALALMRHQPPAELLDSGAAVAAMRASSAGGQNCANTLYALAVLQLRH
AGLEAAVCGRLGELLQHDSDTVLHQHVCNSLWAMAVFGGTYSPDMKQLSIQLAREAVSRWEGLTTEHLKQ
LWQAQQELGGEVAAALGSGSGLQAAMAAAVTTEREDAKPTSDTQKQVVAALRRLEQQGLKAAGGLAVQSL
QTGVVAPGVLAPVDAVVRLSDGQQAAVEFLGPKRYLLTQTQQQDPSAAVEGGTAIRIRQLRRAFKEGGVL
LVPYWEWDFLTPSEQEAYLLRRLQQPAVAVETVTAGAVGGAAAVTTAAPHQPATTTTNSNKGGGGSGGNP
QQQLLVRPARRPEGGGGRRRVGGSKRTVGLRS
```

```
>Chlre_NCL8
     mRNA            complement(join(<815802..815598,815378..>813478))
                     /gene="Chlre_NCL8"
                     /product="predicted NCL protein"
     CDS             complement(join(<815802..815598,815378..813478))
                     /gene="Chlre_NCL8"
                     /product="predicted NCL protein"
                       Sequence gap...//GRGSGGGRAGDGSRASGRSRGSGRDPADDM
TVCRTLEELQAIISQRLSVWEGRKDFITMCAAFNLCGKLESRAGASQTAVARSSIIGMLAPAYLPLVPRI
RNAKNCSIPLWALGKAGLVGGGTEAQLAAALLERLVDPEVLGKAEPQNLAIVMYALGKLKQDQQQRGSGW
DPTSSPHLQTLAAAVASRLRTVAGHGFKPQDVSNSLWALKALGCTGPEYRSAVEALCGEALQRLRKPELA
AAFIPQQLSNILLALEGLQLCSEQAELVVAVAAEDVRRGFDGYIAQDLSNSAWALAKMGFGAGPEAPAEQ
RQWVSAAVDAATQSGAMARATSQAWSNLLYALALMRHQPPAELLDSGAAVAAMRASSAGGQNCANTLYAL
AVLQLRHAGLEAAVCGRLGELLQHDSDTVLHQHVCNSLWAMAVFGGTYSPDMKQLSIQLAREAVSRWEGL
TTEHLKQLWQAQQELGGEVAAALGSGSGLQAAMAAAVTTEREDAKPTSDTQKQVVAALRRLEQQGLKAAG
GLAVQSLQTGVVAPGVLAPVDAVVRLSDGQQAAVEFLGPKRYLLTQTQQQDPSAAVEGGTAIRIRQLRRA
FKEGGVLLVPYWEWDFLTPSEQEAYLLRRLQQPAVAVETVTAGAVGGAAAVTTAAPHQPATTTTNSNKGG
GGSGGNPQQQLLARPARRPEGGDGGGGGGGGSKRTVVGLRS
```

```
>Chlre_NCL21
     mRNA            complement(join(<1029175..10284254,1028148..>1025386))
                     /gene="Chlre_NCL21"
                     /product="predicted NCL protein"
     CDS             complement(join(1029175..10284254,1028148..1025386))
                     /gene="Chlre_NCL21"
                     /product="predicted NCL protein"
MLQFAHGSAVGLRRSYNAPASAHIGLELCPFGIRLTQPRAARTAGRLGPSATAGGVARGLACSSATSGPA
AAVLSDGGPLLPATAPAATGAGGGLRQLLPLAAWARQAAPAPATAAPAAAGARSCLLQPGRRAQQQVLAA
AGRYGSGGTIDRSSGRSSSSGHGGIGSSRSSSDAGGRAGGGRAGGGRWAGGRTGGAGDRSGGGRRARDPA
DDMTVCRTLEQLQGVIDQRLAVWCERQDVIAMSAAFGRCGKLESARAGGPAATAAARAGIIAALAPALLP
LVPRIRQPAGCSLPLWALAKAGAA*AASDGRVESQLAPALLQRLVDPVLLDSATPQALANALYALGKLRE
DQQRRGSGWDPTSSPHLKALASAVASRLRAARGHGFNAQDVSNTLWACAKLGYRDSDLLRPLAEAAAALA
PDMNAQGLANSLWALEVLGCTGSAFRAGLEAMYGAALRRLRTPKEAEAFKPQDLSNILLALEGLQLGGKQ
SELLAAAVAAECVRRLRAPKEAEAFKPQELSNILLALEGLQLGGKQSELLAAAVAAECVRRLRAPKEAEA
FKPQELSNILLALEGLQLGGKQSERLAAAVAAEGVRRGFAGFKPQELSNSAWALAKMGYGAGATPQAMEQ
SHWYAAAVAAAERPGVMAGAKPQAWANLLYALALVRHQPPPTLLDAGAAAAMQRGTAQHCANTLWALAVL
QLRHAGLEAAVCGRLGELLRLGPESLVEQALCNSLWALAALAGGGVPASPAAAALAPALAREAVHRIRDL
AGDDFRQLWQAQHELGGEVAEALARSPDLLAAMEAAVAAQRSTTSTTSSTQKQVAEALQRLLQKGRLPIV
SVRTEVVVEGVLGRVDIVADWSDGRRVAIEVDGPDHFLTNRKDDPSAVDGSTALRNRQLRRALGEGGLLC
VPYWEWYERQAPSAQEAYLLQQLQDLLSGASSGASAGEGSAAARRRKEQRTTQPQPYTARSPTSASTTAS
ITTTATTTTTASSSQKRRILVVRRKAPKQEAVGGGGAGGASSATAAAATIPAEAKVVAPALQQQAEESHQ
LPSSGASGGAGEGGGDGASLALQPPPPQQQHVSAPRARRSSSQSRSRSSQAGSPSAPPPPPFEAPPPEQV
QAAAPPRRAAGGKRTERKAAAAPAAGGGGGGGGGSDSGAVVVPAAVSEKR
```

>NCL30

```
     mRNA            join(<1086885..1087661,1087978..>1090275)
                     /gene="Chlre_NCL30"
                     /product="predicted NCL protein"
     CDS             join(1086885..1087661,1087978..1090275)
                     /gene="Chlre_NCL30"
                     /product="predicted NCL protein"
MLAQLSRALILASATGARGQALGSLSSAIRNRCLEHHQHLTTTTEDGCVVRSAPLGSRCVGLS
HPAAAAGSEAPAAADHSSDLLQLPLPLPASHLAYTTSYSSVISGPAAAGLSGGGPRLAAAAAP
AMFTPGAAAAAAYGSLPPPLGWHAQPQVAAAGVMHGSGTTGSASSGSSGSSSSSSSSGEGDGG
NGRSSGGGRGSGRTGGNGGGGGGRARDPTDDMTVCRTLEDLQAVIDQRLAVWCERQDVSAMSA
ALGRCGKLESARAGPAATAAARAGIIAALAPALLPLVPRISKPVDCSIPLWALAKAGAASDGR
VESQLAPALLQRLVDPVLLDSATPQNLANALYALGKLREDQQQRGSGWDPTSSPHLIALTGAV
ASRLRAARGHGFTAQNVSNSLWACAKLGYRNSAILLPLAEAAAALAPDMNAQELANSLWALEA
LGCTGPAFRAVLETLYGAALRRLRAPKEAEAFKLQGLSNILLALEGLKLRGKQSRLLAAAVAA
ECVRRGFAGFKPQHLSNSAWALAKMGYGAGATLHAAEQRHWYAAAVAAAQRPGVMAGAKPQAW
ANLLYALALVRHQPPPALLDAGAAAAMQRGNAQACANTLWALAVLQLRHAGLEAAVCGRLGEL
LRQGPESLVAQDLCNSLWAVAVLAGGGGPASPAASALAPALAREAVRRRQEFQTEGLLQLWQA
RQELGGEVAEALARSPDLLAAMEAAVAAERATGLNTSRLQEQVAEALQRLLQKGRLPIVSVRT
EVVVEGVLGRVDIVADWSDGRRVAIEVDGPDHFLTNRKDDPSAVDGSTALRNRQLRRAFGEGG
LLCVPYWDWCGRKTPSFQEAYLLQQLQDLLSGAPSGAAAGGEGSAAAPRRKQQRTTQPQPDTA
GTTTTATTTASSSQKRRILVVRRKAPKQDLQHDATCDAPSASGAGPAEAVAEVATSGGPRPAT
AGADVASAPRRQQRQRQLQQQPTRPEPEAVAAAVGRRRRPAAKEPRPQQQQTRPPPQNTRKRA
ASANAHAAAAAAAAAA
```

>Chlre_NCL35

```
     mRNA            complement(join(<5869235..5869206,
                     5868844..5868225,5868126..>5866377))
                     /gene="Chlre_NCL35"
                     /product="predicted NCL protein"
     CDS             complement(join(5869235..5869206,
                     5868844..5868225,5868126..5866377))
                     /gene="Chlre_NCL35"
                     /product="predicted NCL protein"
MSAAFNLCGKLESARAGGPAATAAARAGIMAALAPALLPLVPRIRQPAGCSIPLWALAKAGAASDGRVAS
QLAPALLQRLLDPVLLEAAKPQELSNALYALGKLREDHQQRGSGWDPTSSPYLNALAGAVASRLRAAEGH
GFKPQELSNSLRACAKLGYRDSALLLLLSEAAAALAPDMSAQDLANSLWALDALGSTGPEFRAVLETLCG
AALRQLRTPKEAEAFKPQGLSNILLALEGLQLGGQQSERLAAAVAAEGVQRGFAGFNPQALSNSAWALAK
MGYGAGATPQATEQSHWYAAAVAAAERPRVMAGAKPQEWANLLYALALVRHQPPPALLDAGAVVAMQRGN
AQECANTLWALAVLQLRHAGLEAAVCGRLGELLRLGPESLVAQNLCNSVWALAVLAGGGVPASPAAAALA
PALAREAVRRREQFVDMSAEHLRQLWQAQHELGGEMAEALARSPDLNAAMEAAVAAERATGSNTSSTQKQ
VAKALQRLLQKDLLPIVSLQTEPVVEGVLGRVDIVADWSDGRRVAIEVDGPDHFLTTRKDDPYAVIGSTA
LRNRQLRRAFGRGGLLCVPYWEWDRRKTPSAQEAYMLQQLQDLLSGASSGAAAGEGSAATPRRKQQRTIP
PQPDTAGSSSITTATTSTTASSSQKRRVLVVRRTAPKQEAVGTSGGAGGASDAAAAATATAAAAGAKSVA
PAWQQQAEEAQRLPSSGGAGEGGGDGAPLAPQPPPPQQPRVSAPPTRSSSRQS
```

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
CGCACCTGCAGCAGCCGGGGCGCGCAGCTGTCTCCTGCAGCCGGGGCGGCGCGCAGCAGCAGGTGCTG
GCGGCTGCAGGGAGGTACGGCAGTGGCGGCACCAGTCACAGGAGCAGTGGCAGGAGCAGTGACAACAGCA
GCAGCGGACATGGCGGCATCAGCGGCAGCCGCAGTAGCAGTGACGGTGGCGGAAGGGCGGGCAGGGGCAG
GGCAGACGGAGGCCATTGGGCCGGCGGCAGCGGGCGTGGCAGTGGAAGGACGGGCGGTGGTGGTGGCAGG
AGTGGTGGTGGCGGTAGGCCGCGTGACCCCGAAGACGTGATGATCGGGTATGCGACTCTGGAGGAGCTGC
AAGGAGTCATTGATCAGCGGCTGGCGGTGTGGTGCGAGCGGCAAGACGTGAGCACCATGTCAGCTGCCTT
CGGCCGTTGTGGCAAGGTGGGAGGTTGGACTAGCTAGCGGGCAATAAGAAGAGGAGGGGCAGGGGCAGGG
GTCGGGCTTGCCTGCAAGCACGTGCGGGCCTTCCTGCATTGCAGTGCGGGCACTGACAACGCCAACAACA
CACGCACCTCCTTAGCACTTACCAAAGCTCACACGGTACTGCTTAGGGATGCGTTGCATGATGACGGTGG
GGCTGATCCAAGCTTGCTTTTCACCGCCACAGTGAGTGATCGGGGCATCGTCCTTTCCCACTGCCCCCCC
GCCCCCCCGCCCCCCCCCCTGCAGCTAGAGTCAGCACGCGCCGGCGGGGCGACCGCCACAGCCGCGGCCC
GTGCCAACATCATTGCGGCCCTTGCGCCTGCTCTCCTTCCGCTCGTCCCACGCATCCGGCAGCCGGCCGG
CTGCAGCCTTCCGCTGTTGGCTCTGGCGAAGGCCGGGGCTGCCAGCGACGGGCGTGTGGAGTCACAGCTG
GCGCCGGCACTGCTGCAGCGGCTGGTGGATCCGGTGTTGCTGGATAGTGCCACTCCCCAGGCCCTGGCGA
ACTCACTGTATGCGCTGGGCAAGCTGCGGGAGGATCAGCAGCGGCGGGGCTTGGGCTGGGATCCCACCTC
ATCGCCGCATCTGAAAGCGCTGGCGAGCGCTGTGGCCTCGCGGCTGCGTGCGGCGAGGGGGCACGGCTTC
ACCGCCCAGGGTGTTTCTAACAGCCTGTGGGCGTGCGCCAAGCTGGGGTACCGCGACTCAGCACTCCTGC
TGCCGCTGGCGGAGGCGGCCGCAGCCCTTGCCCCGGACATGAATGCGCAAGACCTGGCCAACAGCCTGTG
GGCGCTGGAGGCTTTGGGCTGCACCAGACCGGAGTTTCGGGCAGTGCTTAAGGCGCTGTGCGGTGCGGCG
CTGCGGCAGCTGCGAACCCCGAAGGAAGCTGCAGCGTTCATTCCGAGACACCTGTCCAACATCTTGCTGG
CGCTGGAGGGGCTGCAGCTGCGAGGCAAGCAGTCCGAGCTGCTGGCGTCGGCTGTGGCTGCAGAGGGTGT
GCGGCGGGGCTTTGCAGGCTTTAACCCGCAAGACCTCAGCAACTCGGCGTGGGCACTGGCCAAGATGGGG
TATGGGGCTGGTGCAACGCCGCAAGCCATGGAGCAGAGCCATTGGTATGCGGCCGCGGTGGCAGCAGCTC
AGCGGCCCGGGGTCATGGCCGGCGCCAAGCCGCAGGCCTGGGCCAACCTGCTGTACGCGCTTGCGTTGGT
GCGCCACCAGCCACCGCCCACGCTGCTTGACGCAGGCGCCGCGGCGGCGATGCAGGGAGGCAATGCACAG
GAATGTTCAAACACGCTGTGGGCGCTGGCGTTGCTGCAGCTGCGGCATGCCGGCTTGGAGGCGGCGGTGT
GCGGCCGGCTGGGTGAGCTGCTGCGGCAAGGCCCGGAGTCGCTTATTGCACAGGAGCTTTGCAACAGCCT
GTGGGCGCTGGCAGTGCTGGCAGGCGGCGGTGGCCCTGCTAGCCCGGCTGCCGCTGCGATGGCCCCGGCG
CTTGCCCGCGAAGCCGCGCGGCGCCGTGAGGGGTTTCAGACCGAGGGGCTACTGCAGCTCTGGCAAGCGC
AACATGAGCTGGGCGGCGAGGTGGCGGAGGCTCTTGCCCGCAGCCCCGACCTGCTAGCCGCGATTGAGGC
GGCGGTAGCTGCCGTACAGGCGACGGGGTCGACCACCAGCAGTCTGCAGGAGCAGGTGGCAGAGGCGCTT
CAGCGACTGCTGCAGAAGGGTCGCCTGCCCATCGTTTCGGTGCGGACGGAGGTGGTGGTGGAGGGCGTGC
TGGGGCGTGTGGATATTGTGGCGGACTGGAGCGATGGGCGCAGAGTGGCAATCGAGGTGGACGGGCCAGA
CCACTTCCTCACCAACCGGAAGGACGACCCGTCCGCGGTAAACGGATCCACGGCCCTACGTAACCGGCAG
CTGCAGCGGGCGCTAGGCGAGGGAGGCCTGGTGTGTGTGCCGTTCTGGGAGTGGTACGGTCTGAGGACGC
CTTCCGCTCAGGAGGCCTACCTGCTGCAGCGGCTGCAGGATCTGCTATCGGGCGCTTCGTCAGGAGCAGC
AGGAGGGGGAGGCAGCGCCGCCGCCCCTCGCCACCAGCAGCAGCGGACGACTCGGCTGCAGCCGGACATT
GCCGGCAGCAGCACCACCAGCACCACCAGCACCGCTAACAGCAGCCAGAAGAGGCGCGTCCTCGTTGTCC
GCAGAAAGGCGCTGAAGCAGGAGGCGGCCGGCGGCGGTGGGGCGGGAGGGGCCTCGAATGCAACAGCAGC
TGCAGCGACGGTAGCGGTTGCGGGGGCTGAGTCAGTGGCCCCTGCCTGGCAGCAACCAGTGGAAAAGGCG
CAGCGGCTTCCCAGCAGTGGCGCTGGTGGCGGTGCTGGCGAGGGCGGGGAAGACGGAGCTTCTCTGGCGC
CGCAGCCTCCACCACCGCAGCAGCAGCATTGTCAGCGCCCCGGGCGCGCCGCAGCAGCAGCCAAACCCA
AAGCCGTAGCCGCAGTAGCGTGTATACATACCCCTACGACGTTCCGGACTACGCGTACCCGTACGATGTC
CCCGACTACGCTAGCTACCCTTATGATGTTCCTGATTATGCTTGAAGATCT

ncc2-HA

GAATTCATGCGCACCAGCCGGCCGCCTCTCAGCGCGTACAGCAACAGCAGGCATGGCTACGGCTATGGAT
CTGCAGCTGTCAGCCAGGGACCGCGCCCGGCGTCACGCGGGCTGCTGCCTGTTCGCTCGGGCACTACGGT
GCCCAGGGCGTTTGCGAGCCCGAGCTCGGCCTACTCGGCCCGACCTGTCCTGCTGGAGTCGCGCAGCCTC
TCCCGCACGGAGCGGCTCTGTGCCGTCGGTGCTCGCTGGCAGCGGTGGCCGGCGCCCCATCGCCCCCGCCG
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 **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.

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. J Biol Chem **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.