Exploring the potential of yeast mitochondria for synthetic cell research

Koster, C.C.

DOI
10.4233/uuid:ef3e6834-988c-4d61-99f3-94133e005f1d

Publication date
2023

Document Version
Final published version

Citation (APA)
Koster, C. C. (2023). *Exploring the potential of yeast mitochondria for synthetic cell research*. [Dissertation (TU Delft), Delft University of Technology]. https://doi.org/10.4233/uuid:ef3e6834-988c-4d61-99f3-94133e005f1d

Important note
To cite this publication, please use the final published version (if applicable).
Please check the document version above.

EXPLORING THE POTENTIAL OF YEAST MITOCHONDRIA FOR SYNTHETIC CELL RESEARCH

CHARLOTTE C. KOSTER

EXPLORING THE POTENTIAL OF YEAST MITOCHONDRIA FOR SYNTHETIC CELL RESEARCH

Proefschrift

ter verkrijging van de graad van doctor
aan de Technische Universiteit Delft,
op gezag van de Rector Magnificus, prof.dr.ir. T.H.J.J. van der Hagen;
voorzitter van het College voor Promoties,
in het openbaar te verdedigen op
vrijdag 22 september 2023 om 12:30 uur

door

Charlotte Catharina KOSTER

Master of Science in Life Science & Technology, Technische Universiteit Delft, Nederland
geboren te Voorburg, Nederland

Dit proefschrift is goedgekeurd door de promotoren
prof.dr.ir. P.A.S. Daran-Lapujade en prof.dr. J.T. Pronk

Samenstelling promotiecommissie bestaat uit:

Rector Magnificus	Voorzitter
Prof.dr.ir. P.A.S. Daran-Lapujade	Technische Universiteit Delft, promotor
Prof.dr. J.T. Pronk	Technische Universiteit Delft, promotor
Onafhankelijke leden:	
Prof.dr.ir. S.J.J. Brouns	Technische Universiteit Delft
Prof.dr. M.A. Huijnen	Radboud Universiteit
Prof.dr. L.M. Veenhoff	Rijksuniversiteit Groningen
Prof.dr. J. van der Oost	Wageningen University & Research
dr.ir. M.M.M. Bisschops	Wageningen University & Research
Prof.dr. F.J.J. Hollmann	Technische Universiteit Delft, reservelid

The research presented in this thesis was performed at the Industrial Microbiology section, Department of Biotechnology, Faculty of Applied Sciences, Delft University of Technology, the Netherlands, within the research programme BaSyC - Building A Synthetic Cell of the Netherlands Organisation for Scientific Research (NWO).

Keywords: *Mitochondria, RNA import, transcriptomics, arginine, synthetic cell, synthetic biology*

Printing: Ridderprint, www.ridderprint.nl
Design & cover: Charlotte C. Koster

ISBN 978-94-6366-717-3
An electronic copy of this dissertation is available at https://repository.tudelft.nl
This thesis is printed on fully recycled paper.

Students at the Technische Hogeschool Delft, circa 1914. From the Almanak of the D.V.S.V. 1914, personal property of the author.

you are one person
but when you move
an entire community
walks through you

- You go nowhere alone
Home Body, Rupi Kaur

The DNA of mitochondria, 'the powerhouse of the cell', is passed on through the maternal line. Therefore, I would like to dedicate this book to the women who have inspired me, and those who inspire the next generations of (female) scientists, striving for inclusivity and equality in science and society.

TABLE OF CONTENTS

SUMMARY

Over the past thirty years, technologies to precisely edit DNA have dramatically improved, kickstarting the field of synthetic biology. Synthetic biology includes all research involving rational design and re-design of living organisms and their functions, making the engineering of cells more predictable, standardized, and modular. Synthetic biology is mostly applied to enable the engineering of tissues, (micro)organisms or microorganism-derived compounds for use in different applications, often with a focus on human health, ecology, or sustainability. Although synthetic biology traditionally focuses on engineering pre-existing organisms, the field is also branching out in a new direction: building fully synthetic, rationally designed minimal cells. Building synthetic cells is extremely interesting from a fundamental perspective, as the ability to rationally build viable, dividing, and self-maintaining synthetic minimal cells provides knowledge on the minimal requirements to sustain life. This should enhance our understanding about which biological parts are minimally required for a cell to live, and how different cellular functions work and interact with each other. In addition to this, such a minimal cell can provide new insights in origins of cellular life and how this evolved. Besides a fundamental understanding of life, synthetic cells can also be applied as synthetic biology tools, for example for synthesis and delivery of therapeutics, or the production of compounds that cannot be produced with currently available organisms used as cell factories.

The research presented in this thesis was performed within the Building a Synthetic Cell (BaSyC) consortium, that aims to build a synthetic cell completely from scratch. Starting with simple biological parts, such as DNA, RNA, proteins and lipids, the research efforts in BaSyC aim to bring simple biological components together to build a synthetic cell. The research within the consortium is split between different parts of the synthetic cell, and the presented research focused on building the genome of a synthetic cell. Therefore, the overarching goal of the research described in this thesis was to devise a strategy for building the genome of this synthetic cell, using baker's yeast *Saccharomyces cerevisiae.*

Two routes towards building a synthetic genome were explored: either building a genome *de novo* using yeast *in vivo* assembly or engineering and expanding the pre-existing minimal genome of *S. cerevisiae* mitochondria. The introductory **Chapter 1** provides an overview of approaches that are used to design and build synthetic cells and summarizes of the progress and applications of synthetic cells. Additionally, it provides a rationale to how mitochondria can be used as starting point of synthetic cells and lists the challenges that are associated with research into mitochondrial biology and engineering of mitochondrial genomes, including challenges in targeting proteins, DNA and RNA to the mitochondria.

Building a synthetic genome requires a method to assemble DNA, as either separate genes or expression cassettes into a full genome. In **Chapter 2**, the potential of *S. cerevisiae* to act as a 'genome foundry' is explored, within the scope of synthetic genomics. Synthetic genomics is a field that specifically focuses on the rational design and construction of chromosomes and genomes for synthetic cells and cell factories. Currently, progress in synthetic genomics is mostly hindered by the inability to synthesize DNA molecules longer than a few hundred base pairs *in vitro*, while the intended size of the synthetic cell genome would be over one hundred kilobase pairs long. Therefore, a method to assemble these small synthetic fragments of DNA into long synthetic genomes is needed. A widely used approach for the assembly of pieces of DNA is yeast *in vivo* assembly. In this method, linear fragments of DNA with homologous ends are transformed into the yeast. The efficient DNA repair machinery of *S. cerevisiae* will then attach the DNA-ends through a process called homologous recombination. This approach has already been used to assemble DNA up to several hundred kilobase pairs, but its limits have not yet been reached. In **Chapter 2**, the pivotal role of *Saccharomyces cerevisiae* in the establishment of synthetic genomics is reviewed, and an overview is provided of the various genomes that have been assembled using this yeast, while also looking forward to the key role that the *S. cerevisiae* genome foundry will play in the future development of synthetic genomics and synthetic cells.

Nevertheless, the construction of synthetic cells is not advanced to the level where an assembled genome

can readily be encapsulated and expressed. As described in **Chapter 2**, there are still gaps in knowledge on the compatibility of yeast *in vivo* assembly with exogenous or repetitive DNA sequences. To overcome these challenges, we envisioned the engineering and expansion of the preexisting minimal genome of yeast mitochondria as an intermediate for building synthetic cells. Before the mitochondrial genome can be engineered, two challenges were identified. The first is a limited understanding of gene expression of *S. cerevisiae* mitochondria, which was the focus of **Chapter 3**. A second, more technical challenge is the unavailability of DNA editing technologies for the integration of genes in the mitochondrial DNA, which was further explored in **Chapters 4** and **5**.

The mitochondrial genome is expressed as long polycistronic RNA transcripts that undergo post-transcriptional processing and splicing. In **Chapter 3**, we identified a gap in knowledge on the expression and processing of mitochondrial RNA. The mitochondrial transcriptome has a polycistronic, and therefore, complex, mosaic structure. Additionally, the analysis of the mitochondrial transcriptome is complicated, as the transcriptome represents only a small fraction of the total cell transcriptome and does not co-purify with the cytosolic transcriptome. Because of this, the mitochondrial transcriptome is rarely studied in transcriptome analysis and fundamental questions regarding mitochondrial gene expression and splicing remain unresolved, even in the well-studied mitochondria of *S. cerevisiae*.
These challenges could be overcome by long-read RNA-sequencing to elucidate the different RNA structures present in the complex mitochondrial transcriptome, and by enrichment of the mitochondrial transcriptome fraction. To this end, a method for enrichment of mitochondrial RNA and sequencing using isolation of mitochondria and subsequent RNA-sequencing using Nanopore technology was developed, which is described in **Chapter 3**. This method successfully captured the full mitochondrial transcriptome, quantified all transcripts, and resolved RNA splicing patterns with a single base pair resolution. Subsequently, this method was applied to explore the transcriptome response of *S. cerevisiae* grown with glucose or ethanol as sole carbon source, and specifically study the impact of growth conditions on mitochondrial RNA-expression and splicing. The research presented in this chapter uncovered a remarkable difference in abundance of group II introns between yeast grown in mostly fermentative and fully respiratory conditions. Characterization of a strain devoid of introns did not show a profound impact on respiratory activity. Therefore, this chapter opens new questions as to what impact this accumulation of introns has on mitochondrial functions.

The genetic engineering of a genome often requires the use of auxotrophic selectable marker genes. These genes encode a metabolic trait that can be used to select engineered cells by omitting a nutrient in the growth medium. For mitochondrial genome engineering, only one of these auxotrophic marker genes exists: *ARG8*. The *ARG8* gene encodes the mitochondrially localized acetylornithine aminotransferase, which catalyzes an essential step in arginine biosynthesis.
In **Chapter 4**, a frequently occurring revertant mutation was found that could bypass arginine auxotrophy in *Δarg8* mutants, rendering the only auxotrophic selectable marker ineffective. Additionally, the arginine prototrophic *Δarg8* strain was respiratory deficient. Through whole-genome sequencing, a mutation in *UME6* was identified as the underlying molecular mechanism that allowed the mutants to synthesize arginine. Ume6 is a transcriptional regulator involved in many pathways, including arginine synthesis. The mutation most likely caused a loss of function of Ume6, and thereby the de-repression of the *CAR2* gene. Car2 catalyzes a reaction leading towards a precursor of arginine, providing the cells with an alternative route for arginine biosynthesis. Double deletion of *ARG8* and *CAR2* caused complete and stable arginine auxotrophy. An *Δarg8Δcar2 S. cerevisiae* strain can be used for future mitochondrial targeting and expression studies using the *ARG8* marker.

An efficient method to engineer the mitochondrial genome is highly sought after for both therapeutic purposes as well as synthetic biology applications. Despite many efforts, the integration of genes in the mitochondrial genome using the efficient and flexible CRISPR/RNA-guided endonucleases (RGENs) has not yet been achieved. The use of RGENs in mitochondria is limited by a lack of understanding of mitochondrial RNA import, which is essential for implementation of RGEN-mediated genome engineering of mitochondria. Several studies have reported the import of short RNA in mitochondria, leading to several putative import mechanisms and import determinants that may facilitate import of short RNA. However, evidence for RNA

import is often circumstantial as there are no methods to directly determine and quantify the import of RNA without false-positives.
Therefore, **Chapter 5** describes the design and testing of experimental strategies to detect and improve the import of mRNA-sized RNA in mitochondria, using *S. cerevisiae*. A fluorescence-based screening approach was developed to screen for mitochondrial translocation of mRNA containing RNA import signals. Although mitochondrial translocation of mRNA was detected, import appeared to be random and dependent on the co-targeting of another fluorescent protein to the mitochondrial matrix. Conversely, it appeared that the used RNA-import signals did not have any impact on the targeting of RNA to the mitochondria.
Currently, sufficient knowledge to rationally engineer RNA import is lacking. Therefore, **Chapter 5** also describes an evolutionary engineering approach for RNA import based on the *ARG8* marker described in **Chapter 4**. This approach was employed to identify potential mechanisms for the import of RNA and to investigate whether efficient mitochondrial mRNA import could be established. Although evolutionary engineering over 500 generations resulted in limited arginine prototrophy, whole-genome sequencing and proteomics revealed this was due to altered protein expression and turnover, likely resulting in altered resource allocation or recycling of arginine. This analysis confirmed mRNA import in mitochondria was not achieved despite the extensive evolution under a strong selective pressure, highlighting the complexity of mitochondrial RNA import. Nevertheless, this approach is a promising, unambiguous method for future import studies with different (shorter) RNAs or yeast mutants.

SAMENVATTING

In de afgelopen dertig jaar hebben de technieken om DNA op een heel precieze manier te bewerken een verregaande ontwikkeling doorgemaakt. Dit heeft het onderzoeksveld van synthetische biologie tot stand gebracht. Synthetische biologie omvat al het onderzoek naar het rationeel ontwerpen en herontwerpen van levende organismen. Hierdoor kunnen cellen worden aangepast op een voorspelbare, gestandaardiseerde en modulaire manier. Synthetische biologie wordt voornamelijk toegepast in het modificeren van weefsels, (micro)organismen, of materie gemaakt van en door micro-organismen, en vindt voornamelijk toepassingen op het gebied van menselijke gezondheid, ecologie of duurzaamheid. Synthetische biologie is oorspronkelijk ontwikkeld met als doel reeds bestaande organismen te modificeren, maar het veld beweegt zich in een nieuwe richting: het maken van nieuwe, volledig synthetische en rationeel ontworpen minimale cellen. Het bouwen van een synthetische cel brengt veel fundamentele wetenschappelijke kennis met zich mee. Het op een rationele manier opbouwen van een levende, delende en zelf-onderhoudende cel biedt namelijk inzicht in de minimale vereisten die ten grondslag liggen aan het leven. Hierdoor kunnen we beter begrijpen welke biologische onderdelen minimaal nodig zijn om een cel in leven te houden, hoe verschillende onderdelen van een cel werken en hoe deze onderling met elkaar samenwerken. Daarnaast kan een minimale cel inzicht bieden in hoe cellulair leven is ontstaan en is geëvolueerd. Naast een beter fundamenteel begrip van hoe levende cellen werken, kunnen synthetische cellen ook als hulpmiddelen toegepast worden, om bijvoorbeeld geneesmiddelen te produceren en deze op de juiste plek in het lichaam af te leveren, of voor de productie van stoffen die met het huidige scala van industrieel toegepaste (micro-)organismen (zogenaamde 'cellulaire fabrieken') niet geproduceerd kunnen worden.

Het onderzoek dat in dit proefschrift beschreven wordt, is uitgevoerd binnen het Building a Synthetic Cell (BaSyC) consortium. Het doel van BaSyC is om een synthetische cel stapsgewijs op te bouwen, vanuit de biologische 'bouwstenen' DNA, RNA, eiwitten en vetzuren. Het onderzoek binnen het consortium is opgesplitst in het construeren van verschillende synthetische cel-onderdelen, die uiteindelijk later samen komen tot een synthetische cel. Het onderzoek dat beschreven wordt in dit proefschrift richtte zich op het opbouwen van het genoom van de synthetische cel uit DNA. Het overkoepelende doel van dit onderzoek was dan ook om een strategie te bedenken voor het produceren van een synthetisch genoom voor de synthetische cel, waarbij gebruik wordt gemaakt van de bakkersgist *Saccharomyces cerevisiae.*

In dit proefschrift is onderzoek gedaan naar twee verschillende manieren om een synthetisch genoom te bouwen, namelijk het volledig opbouwen van een genoom met behulp van bakkersgist en het uitbreiden van het reeds bestaande, minimale genoom van mitochondriën in gist. In **Hoofdstuk 1** worden verschillende manieren om synthetische cellen te ontwerpen en produceren uiteengezet, en wordt de voortgang van de constructie en de potentiële toepassingen van een synthetische cel samengevat. Verder wordt in dit hoofdstuk beschreven hoe mitochondriën als startpunt van de synthetische cel gebruikt kunnen worden. Tevens wordt er een overzicht gegeven van de uitdagingen die gepaard gaan met het onderzoek naar mitochondriën, waaronder het bewerken van het mitochondriële DNA en het lokaliseren van eiwitten, RNA en DNA in de mitochondriën.

Om een synthetisch genoom te produceren, moet een methode worden ontwikkeld om DNA, in de vorm van losse genen of expressie-cassettes, samen te voegen tot een volledig genoom. In **Hoofdstuk 2** wordt beschreven hoe dit gedaan kan worden met de gist *S. cerevisiae*. Deze gist kan gebruikt worden als een soort 'genoom-fabriek'. In deze discipline, die ook wel synthetische genomica, of '*Synthetic genomics*' heet, wordt onderzoek gedaan naar het rationele ontwerp en de bouw van synthetische chromosomen en genomen van synthetische cellen en cellulaire fabrieken, De ontwikkelingen op het gebied van synthetische genomica worden op dit moment belemmerd doordat het met de huidige technieken niet mogelijk is om DNA-moleculen langer dan een paar honderd baseparen te produceren buiten de cel (*in vitro*). Het theoretisch benodigde genoom voor de synthetische cel is echter minimaal honderdduizend baseparen lang. Een mogelijke oplossing is het ontwikkelen van een methode om korte, *in vitro* gesynthetiseerde, DNA-fragmenten samen te voegen tot een lang synthetisch genoom. Zo'n methode om DNA-fragmenten samen te voegen is al in ontwikkeling, namelijk '*in vivo* assembly' (los vertaald: 'binnen de cel samenvoegen').

De *in vivo* assembly methode berust op het samenvoegen van DNA in de celkern van een micro-organisme. Dit gebeurt door middel van een proces dat homologe recombinatie heet, waarbij gebruik wordt gemaakt van het natuurlijke vermogen van cellen om DNA te repareren. Dit proces is dusdanig efficiënt in gist, dat dit organisme DNA-fragmenten met homologie kan herkennen en aan elkaar kan hechten. Door DNA-fragmenten met overlappende uiteinden in de gistcel te brengen, kan op deze manier een lang stuk DNA gevormd worden uit kleinere DNA-fragmenten. Met deze methode kan DNA van enkele honderdduizenden baseparen lang gemaakt worden, en de maximale lengte van DNA die deze methode kan produceren is nog niet bereikt. In **Hoofdstuk 2** wordt de centrale rol van *Saccharomyces cerevisiae* binnen de synthetische genomica besproken, en wordt een overzicht gegeven van alle genomen die tot nu toe met deze gist zijn geproduceerd. Ook wordt er in dit hoofdstuk vooruit gekeken naar de rol die de 'genoom-fabriek' *S. cerevisiae* zal spelen in toekomstige ontwikkelingen binnen het gebied van synthetische genomen en synthetische cellen.

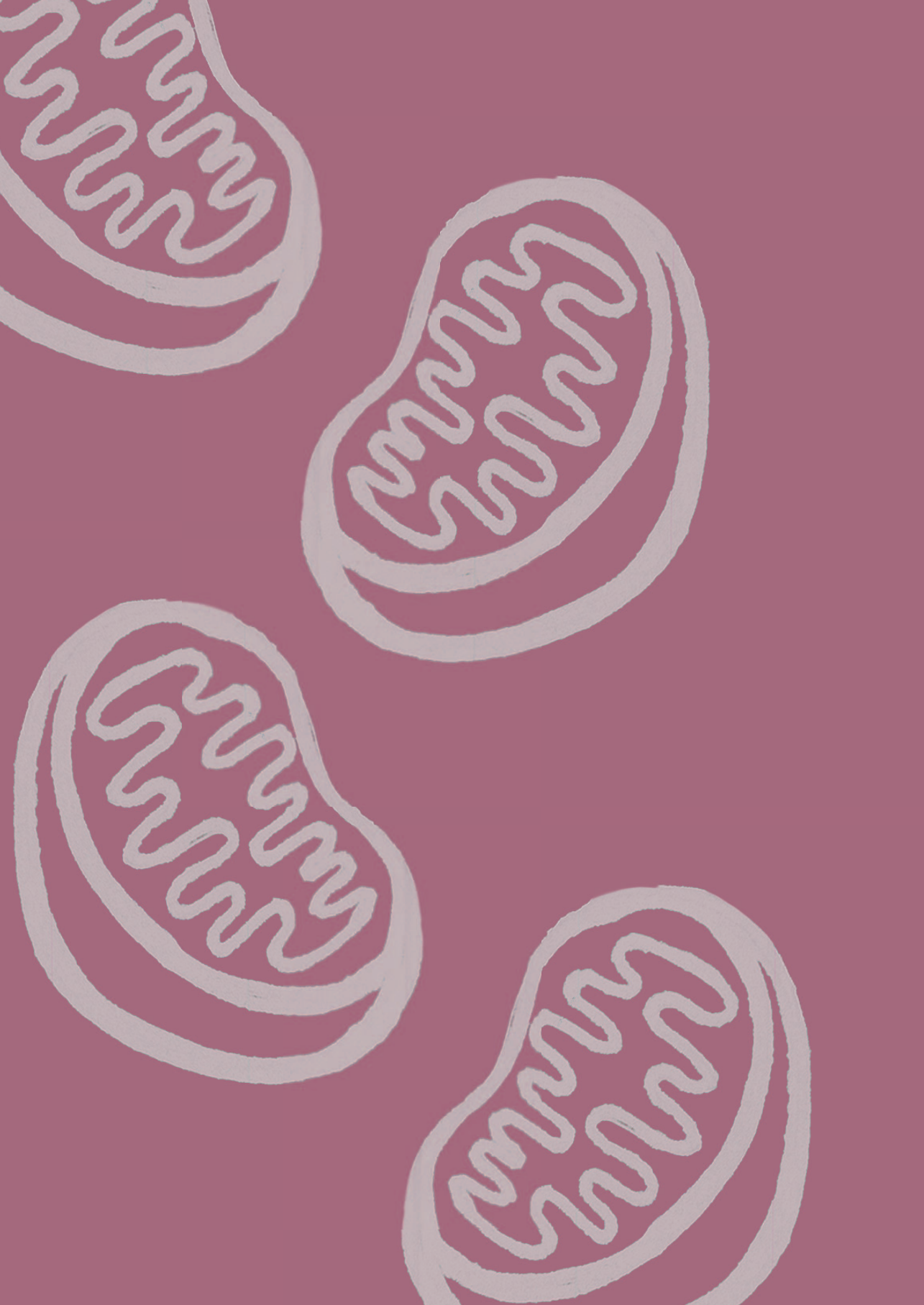
Ondanks deze ontwikkelingen op het gebied van synthetische genomica, is het nog niet mogelijk om een in gist geproduceerd genoom in een synthetische cel te introduceren en tot expressie te brengen, om zo een levende cel te maken. Zoals beschreven in **Hoofdstuk 2**, is er nog veel onbekend over de *in vivo* assembly-methode, waaronder de compatibiliteit met DNA afkomstig van andere organismen dan gist, en met DNA-fragmenten met een repetitieve base-volgorde.
Daarom wordt er in dit proefschrift nog een tweede strategie beschreven om een minimaal synthetisch genoom te produceren. Hierin wordt een genoom niet vanuit het niets opgebouwd, maar wordt het reeds bestaande, zeer minimale, mitochondriële genoom van gist aangepast en uitgebreid als tussenstap voor het maken van een synthetisch genoom. Er zijn echter twee uitdagingen die het aanpassen van het mitochondriële genoom in de weg staan. Ten eerste is er een beter fundamenteel begrip nodig van de manier waarop de mitochondriën van *S. cerevisiae* genen tot expressie brengen, wat uiteen wordt gezet in **Hoofdstuk 3**. De tweede uitdaging is van een meer technische aard: er zijn nu nog geen methodes om nieuwe stukken DNA in het mitochondriële genoom te voegen, dit wordt verder onderzocht in **Hoofdstukken 4** en **5**.

Het mitochondriële genoom van *S. cerevisiae* wordt tot expressie gebracht als lange RNA-moleculen die meerdere genen bevatten en na transcriptie verder worden verwerkt en gesplitst, in een proces dat 'splicing' heet. In **Hoofdstuk 3** worden openstaande vragen rondom de expressie en het verwerken van mitochondrieel RNA, ofwel het transcriptoom, verder onderzocht. Het mitochondriële RNA heeft een complexe structuur, en elk transcript kan op verschillende manieren verwerkt worden. Daarnaast wordt het mitochondriële transcriptoom vaak niet geanalyseerd, omdat dit maar een klein gedeelte van van alle RNA-moleculen in de cel omvat. Wanneer RNA wordt geïsoleerd uit een cel, gaat het mitochondriële RNA vaak verloren. Hierdoor wordt het mitochondriële transcriptoom vaak niet geanalyseerd en is er nog veel onbekend over de expressie en splitsing van mitochondriële genen, zelfs in de goed bestudeerde mitochondriën van *S. cerevisiae*.
Een oplossing hiervoor is het gebruik van nieuwe methodes die lange stukken RNA in één keer kunnen aflezen, zogenaamde 'long-read RNA sequencing' technologieën. Hierdoor kunnen de verschillende RNA-structuren in de mitochondriën in kaart gebracht worden. In **Hoofdstuk 3** wordt de ontwikkeling van een methode beschreven waarin het specifiek isoleren van mitochondrieel RNA wordt gecombineerd met een long-read sequencing techniek voor RNA, die gebruik maakt van de Nanopore-technologie. Met deze methode is het mitochondriële transcriptoom volledig in kaart gebracht. Daarnaast konden de verschillende splitsingspatronen van het mitochondriële RNA tot op de basepaar nauwkeurig in beeld gebracht worden. Deze methode is vervolgens toegepast om te meten wat het effect was op het mitochondriële transcriptoom van *S. cerevisiae* wanneer deze gist glucose of ethanol als enige koolstofbron gebruikte, en specifiek welk effect dit had op de expressie en het splitsingspatroon van het mitochondriële RNA. Het onderzoek in dit hoofdstuk heeft een opzienbarend verschil aangetoond tussen voornamelijk fermenterende gisten op glucose, en volledig respirerende gisten op ethanol, in de hoeveelheid van een specifiek type 'groep II' intronen, die normaliter uit het RNA worden verwijderd. Er was echter geen aantoonbaar verschil in respiratie tussen giststammen met en zonder intron-RNA, dus dit roept nieuwe vragen op over de invloed van de hoeveelheid intron-RNA op de mitochondriën.

Voor het modificeren van een genoom is het gebruik van zogenaamde auxotrofe selectie-genen nodig. Deze genen coderen voor een enzym, dat een essentiële stap in de aanmaak van een cellulaire bouwsteen

katalyseert. Hierdoor kunnen cellen, waarvan het DNA gemodificeerd is met dit gen, worden geselecteerd door ze te groeien op voedingsbodem waarin de desbetreffende bouwsteen is weggelaten. Dit zorgt voor een eenvoudige selectie van genetisch gemodificeerde cellen. Voor de genetische modificatie van het mitochondriële genoom bestaat er slechts één auxotroof selectiegen, *ARG8*. Het *ARG8*-gen codeert voor het mitochondriële eiwit acetylornithine aminotransferase, welke een essentiële stap in de aanmaak van het aminozuur arginine katalyseert.
In **Hoofdstuk 4** is een veel voorkomende mutatie ontdekt in een giststam waarin *ARG8* verwijderd was. Door deze mutatie kon de gist op een andere manier arginine aanmaken, waardoor het enige beschikbare mitochondriële auxotrofe selectie-gen onbruikbaar werd. Bovendien waren deze gistmutanten niet meer in staat om energie te genereren in de ademhalingsketen. In de DNA-sequentie van de mutanten werd een mutatie ontdekt in het gen *UME6*. Het genproduct van *UME6* reguleert de transcriptie van vele metabole routes, waaronder de synthese van arginine. Waarschijnlijk leidde de mutatie tot een non-functioneel Ume6 eiwit, met als gevolg de expressie van het gen *CAR2*, dat normaliter wordt gereguleerd door Ume6. Het genproduct van *CAR2* katalyseert een reactie die leidt tot een grondstof van arginine, waardoor de cellen een alternatieve manier hadden om arginine te produceren. Deletie van zowel *ARG8* als *CAR2* uit het genoom leidde uiteindelijk tot een stabiele arginine auxotrofie. Deze *S. cerevisiae* stam kan gebruikt worden in toekomstige onderzoeken waarin mitochondriële lokalisatie en expressie wordt bestudeerd met behulp van het *ARG8*-gen.

Modificatie van het mitochondriële genoom is niet alleen wenselijk vanuit de synthetische biologie, maar ook vanuit een medisch perspectief. Ondanks vele pogingen is men er nog steeds niet in geslaagd om een nieuw gen in het mitochondriële genoom te integreren, zelfs niet door middel van het doorgaans efficiënte en flexibele CRISPR/RNA-gestuurde endonuclease (RGEN) technologie. Een gebrek aan kennis over mitochondrieel RNA-import, wat essentieel is voor het correct functioneren van een RGEN in de mitochondriën, staat het gebruik van mitochondriële RGENs in de weg. Verscheidene studies hebben de import van korte RNA-moleculen in mitochondriën onderzocht. Dit heeft geleid tot een aantal vermeende mechanismen en factoren die de import van RNA in de mitochondriën zouden faciliteren. Het bewijs voor zulke RNA-import mechanismen is echter vaak indirect, aangezien er geen methodes bestaan om directe RNA-import te meten, wat kan resulteren in foutpositieve resultaten.
Daarom zijn in **Hoofdstuk 5** experimentele strategieën ontworpen en getest om de import van mRNA-moleculen aan te tonen in *S. cerevisiae*. Allereerst werd er een methode ontwikkeld en getest om mRNA import en hypothetische RNA-import factoren te evalueren door middel van fluorescentie. Hoewel met deze methode de verplaatsing van mRNA naar de mitochondriën werd gedetecteerd, leek het importmechanisme willekeurig en leek RNA-import af te hangen van het tot expressie brengen van een ogenschijnlijk niet gerelateerd rood-fluorescerend eiwit in de mitochondriën. Daarentegen leken de eerder beschreven RNA-import factoren geen invloed te hebben op de mate van RNA-import in de mitochondriën.
Er is onvoldoende kennis van RNA-import mechanismen in mitochondriën om hier op een rationele manier een oplossing voor te ontwerpen. Daarom werd in **Hoofdstuk 5** een tweede methode ontwikkeld waarbij een selectiedruk werd uitgeoefend op de cellen die zou kunnen leiden tot evolutie van een RNA-import mechanisme. Hierbij werd gebruik gemaakt van het *ARG8*-gen beschreven in **Hoofdstuk 4**. Door middel van laboratorium-evolutie werd gekeken of RNA werd geïmporteerd in de mitochondriën en of de mogelijke RNA-import mechanismen over tijd verbeterd konden worden. Na meer dan 500 generaties waren de gisten in staat een zeer gelimiteerde hoeveelheid arginine aan te maken. Het aflezen van DNA en mitochondriële eiwitten van de geëvolueerde cellen toonde aan dat dit echter niet kwam door RNA-import, maar door een verandering in het verbruik van arginine in de cel. Deze analyse bevestigde dat de import van RNA niet kon worden aangetoond of verbeterd, ondanks de zeer lange evolutie onder een sterke selectiedruk, en toont aan hoe complex de import van RNA in mitochondriën is. De methode is echter veelbelovend om de import van RNA in de mitochondriën in de toekomst mogelijk te maken, eventueel met andere of kortere RNA's, of kan gebruikt worden om mutanten te vinden die RNA in de mitochondriën kunnen importeren.



CHAPTER 1
INTRODUCTION

1.1 Approaches for building synthetic cells

The role of synthetic biology in society

The ability of humans to precisely edit DNA, the 'blueprint' of all living systems, has rapidly increased over the past three decades, spawning the field of synthetic biology [1]. Synthetic biology is an umbrella term that includes all research involving rational design and re-design of living organisms and their functions. This is preferably done in a standardized and modular way so engineering of microorganisms becomes more predictable [2-4]. Up until now, synthetic biology has focused on building genomes of cells and cell-derived systems, with applications mostly in therapeutics, industry and manufacturing and sustainable agriculture. The host used in synthetic biology efforts varies depending on the application. Synthetic biology in healthcare often focuses on engineering mammalian tissues and developing relevant diagnostics and therapeutics, but also includes engineering of organisms in human microbiomes and production of medicine and vaccines using microorganisms or cell-free systems [5-8]. Industrially applied synthetic biology has a strong focus on fermentation and microbial production of relevant (fine) chemicals, fermented food, and energy sources using engineered microorganisms, often with a strong application in sustainability [9-13]. Synthetic biology in agriculture is often focused on the engineering and improvement of plants and food crops and the newer field of cellular agriculture, or 'cultured meat', but also includes more ecological applications such as plant- or microbial CO_2 and nitrogen-fixation [14-18].

From synthetic biology towards synthetic cells

While the list of organisms 'unlocked' by synthetic biology continues to grow, the field is also branching out in a new direction: building new artificial cell types in the form of fully synthetic, rationally designed

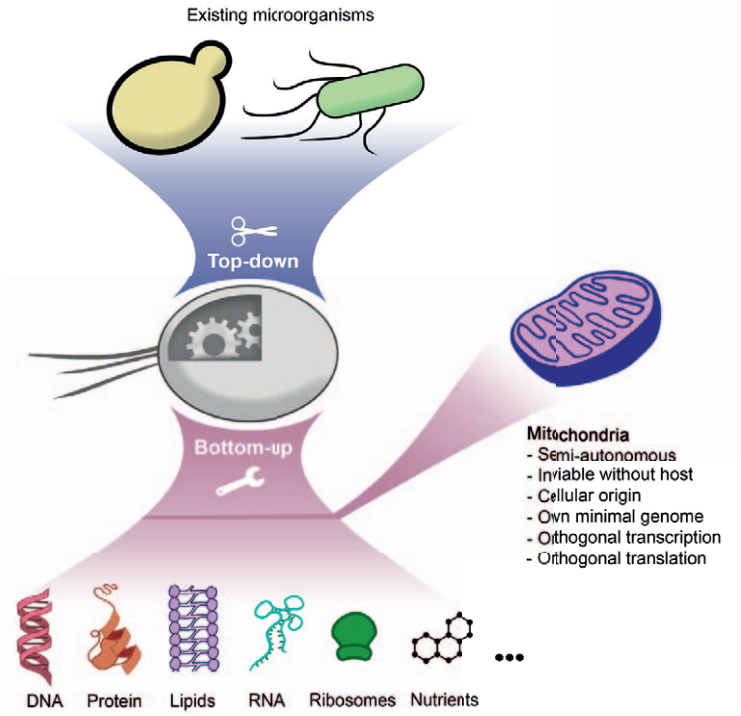


Figure 1.1. Strategies for construction of synthetic cells. In a top-down approach existing microbes are used as a starting point of a synthetic cell, and inessential functions are removed. In bottom-up approaches, biological molecules are combined as building blocks for engineering of synthetic cells. Mitochondria can be considered an intermediate between complete bottom-up biology and a full autonomous synthetic cell.

cells. Building synthetic cells is extremely interesting from a fundamental perspective, as the path towards rationally building viable, dividing, and self-sustaining synthetic cells will generate knowledge on which biological parts are necessary for a cell to divide and maintain life. This will enhance our understanding about the minimal requirements of life, how different cellular functions work and interact with each other, and can provide new insights in origins of cellular life and how this evolved [19-21]. The first big milestone in synthetic cell research would not be to build anything like a prokaryotic or eukaryotic cell, but will consist of a much simpler system that can at least replicate autonomously while mimicking the general makeup of any living cell. This minimal synthetic cell would consist of a lipid membrane that encapsulates DNA that can be transcribed into RNA, which can be translated into protein. The cell would also need a metabolism to provide required building blocks for cell division and an energy metabolism that supports the basal functions of the synthetic cell [22-24]. Upon reaching these objectives, more complex functions such as cell signaling or a complete metabolism can be added. On top of that, bacterial- or even eukaryote-like systems can be built [19, 21, 25]. However, the field is still at the level of curating, designing, and integrating the minimal cellular functions, although many promising efforts have been made, such as *in vitro* DNA replication [26], restriction and membrane growth of liposomes [27, 28], artificial energy and cofactor generation [29-31], motility [32], cytoskeleton reconstitution [33], regeneration of proteins [34] and encapsulation of these modules [35]. Nevertheless, the ambition of most research efforts in synthetic cell biology is to have constructed the first living artificial cell within this decade [36].

Once successful, these synthetic cells may play roles besides fundamental understanding of life, for example as carriers for the synthesis and delivery of therapeutics [37-39]. Targeted and personalized medicine is increasingly more relevant in therapeutics, and although liposome-encapsulated drugs and proteins show promising results, their function could be extended by using synthetic cells. This could include drug delivery, or for in *in situ* biosensors, drug biosynthesis and even has prospects as artificial blood [38, 40-42]. Synthetic cells also may have implications as synthetic cell factories in an industrial setting. Although the current list of available and genetically accessible cell factories and therapeutic microorganisms is rapidly expanding due to the development of new genetic engineering tools and cell-free systems, these applications often require specific microorganisms that are not always able to reach a high product titer or are not always genetically accessible. These technological hurdles can be overcome by using a clean background strain, i.e., a cell that can divide and maintain viability under industrially relevant conditions. These cells can theoretically be fully 'programmed' by adding the required metabolic pathways or cellular functions in the clean background strain, while directing cellular energy to product formation only [43-45]. Synthetic cell research can be divided in two main branches: development of cells that are synthetic, but not necessarily minimal, and the development of minimal synthetic cells. The latter would be optimal for applications in therapeutics or as clean background strain.

Top-down engineering of synthetic cells

A synthetic cell could either be a re-designed or new cell type that is unlike any living form of cellular life, or it could be a minimal synthetic cell, which are synthetic cells that only contain the most basal requirements to sustain life. Historically, the first synthetic cells were generated through a 'top-down' approach. The top-down approach builds on pre-existing organisms: by systematic removal or re-shuffling of genes, a reduced and well-understood set of genes is curated. The minimal set of genes is then re-assembled into a synthetic genome and can be transplanted in a host to generate a synthetic cell with a smaller, re-engineered modular or better understood genome (Figure 1.1) [46-48]. These synthetic cells also may express genes through orthogonal expression systems that use non-native promoters and RNA polymerases, express novel metabolic pathways, or utilize non-natural amino acids and tRNA to alter codon usage [49-51]. Although the top-down research into synthetic cells is widespread in labs all over the world, the field has a strong focus on building synthetic genomes (synthetic genomics, reviewed in Chapter 2) and is mostly performed by a few leading institutes or consortia. Bacterial, algal, viral and mammalian synthetic genomics is mostly performed within the J. Craig Venter Institute (JCVI) while the Sc2.0 consortium explores synthetic yeast genomes [52-54], and leading work on recoding and orthogonality is the expertise of the labs of Jason Chin and George Church [54-56]. The top-down approach has already successfully yielded multiple reduced or re-engineered genomes and artificial cell types, mostly of bacterial and yeast origin [48, 55, 57-60]. However, this may not

be a fruitful way to generate minimal synthetic cells. The most minimal artificial cell to date achieved through a top-down approach, JCVI-Syn3A, contains only 452 essential genes [61]. Even in this minimal synthetic cell, one-third of the genes remain of unknown function, highlighting the difficulty that is involved in curating a minimally required set of genes top-down. This is often complicated by conditional essentiality of genes [62, 63]. Therefore, this approach may yield useful synthetic chassis organisms for several applications [54], but may not provide sufficient insight in the minimal requirements of life.

Bottom-up engineering of synthetic cells

An alternative way of generating synthetic cells is the bottom-up approach, in which a set of biological or even synthetic parts of cells are rationally curated and combined to form a synthetic cell [19, 21]. Such a bottom-up approach constitutes the integration of structural parts of a cell, such as the cell membrane, a cytoskeleton and cell growth and division machineries, as well as functional parts including, but not limited to, metabolism, protein expression machineries, and genome maintenance and division (Figure 1.1) [21, 24, 64]. Since parts are rationally combined, this approach is expected to yield minimal and well-understood living cells. Building a cell from the bottom-up is the scope of the Dutch research consortium Building a Synthetic Cell (BaSyC, www.basyc.nl), of which the research presented in this thesis was part of. BaSyC is part of a larger worldwide community of bottom-up synthetic cell research, including the German MaxSynBio, UK-based BrisSynbio and FabriCell, the European Synthetic cell initiative and the US-based Build-a-cell initiative [64-68].

Even though theoretically the rational design of a synthetic cell would be more suited for the design of minimal cells than the top-down approach, the field of bottom-up biology is still in a developmental phase, and several fundamental challenges need to be overcome before a synthetic cell can be built. This includes identifying which components are required for the assembly of a synthetic cell, how to reconstitute those *in vitro*, how these components will eventually self-organize, how the synthetic cell will be 'booted up' upon assembly, and, last but not least, how to synthetize the synthetic genome of these cells [20, 69, 70]. The genome will be the core of any synthetic cell, encoding all the genes required for the structure and functions of the cell, which would hypothetically encode between 100 to 200 genes and be approximately 100-200 kb in length [71, 72]. A genome of such size could be assembled with a variety of techniques, including *in vivo* assembly in yeast [52, 73, 74]. However, despite the 'awesome power of yeast genetics' to efficiently assemble DNA, this method is still under development for the assembly of long, potentially repetitive, DNA sequences (reviewed in Chapter 2 of this thesis, [70]). To escape some of the challenges faced by the bottom-up construction of minimal cells, this thesis explored a third approach towards building synthetic cells: using the minimal and semi-autonomous mitochondria of *Saccharomyces cerevisiae* as an intermediate starting point (between bottom-up and top-down) for synthetic cell construction.

Mitochondria: a type of minimal cell

Mitochondria are a special type of organelle, as they have a cellular origin. The current consensus is that mitochondria occurred from an endosymbiotic event, where an α-proteobacterial cell integrated inside an archaeal cell. This led to a unique endosymbiotic relationship in which the α-proteobacterial mitochondrial ancestor transitioned from an independent cell into an organelle that was dependent on its archaeal host, resulting in modern eukaryotic cells [75, 76]. The only other instance in which such a relation is found are plant chloroplasts [77]. In this transition, a majority of the mitochondrial ancestor's genome was transferred to the nuclear genome of its host, while the mitochondrial ancestor evolved protein machinery to import the proteins encoded by its host. The metabolisms of the two cells became co-dependent on each other, where the mitochondrial ancestor provided metabolic energy and cofactors required for the archaeal proteins and pathways to function [78-80]. This resulted in a unique host-organelle relation, in which the mitochondria still contain their own minimal genome and the tools to translate this genome as well as a functional membrane and energy metabolism, while the expression of proteins involved in all other mitochondrial processes and mitochondrial maintenance is controlled by the host cell [81-83].

This study focuses on the use of mitochondria of the yeast *Saccharomyces cerevisiae* as a starting point between top-down and bottom-up approaches for building synthetic cells. The mitochondrial genome and its functions are minimal and relatively well-understood [83-85], while mitochondrial morphology and

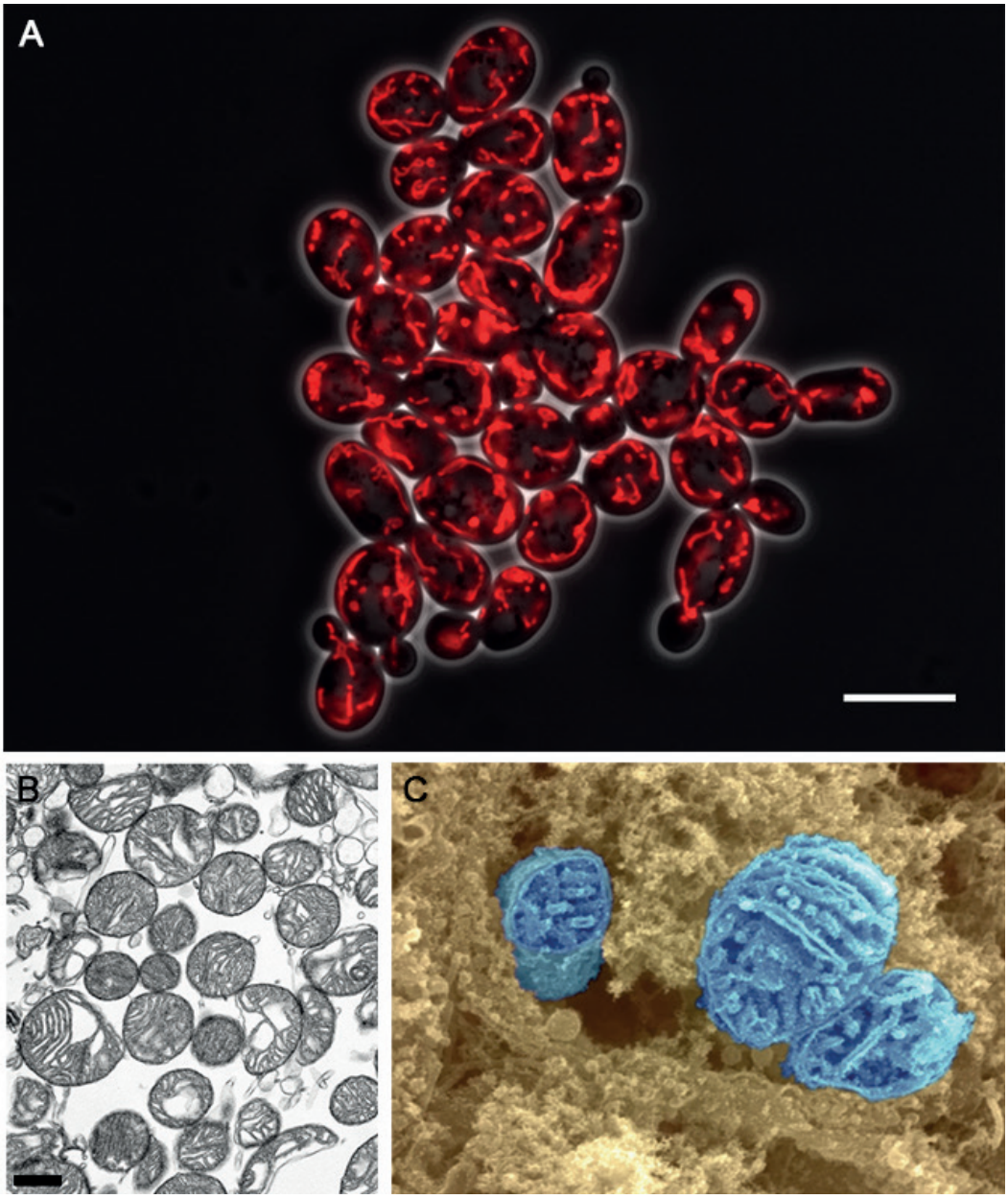


Figure 1.2. The morphology of mitochondria. A) *S. cerevisiae* cells with red fluorescent mitochondria imaged using light microscopy. The scale bar represents 5 µm. Photo credit: Kavish Kohabir. B) Isolated mammalian muscle mitochondria imaged using transmission electron microscopy, showing the cristae structure of the inner membrane. The scale bar represents 0.5 µm. Image adapted from Lai *et al.* (2018) [104] © Scandinavian Physiological Society. Published by John Wiley & Sons Ltd. C) Scanning electron micrograph of a freeze-fractured sample of three mitochondria surrounded by cytoplasm. Image obtained by Dr. David Furness. Attribution-NonCommercial 4.0 International (CC BY-NC 4.0)

maintenance is controlled by the yeast nuclear genome [86, 87]. By expressing modules of the synthetic cell from the mitochondrial genome, the challenges presented by self-assembly and 'kick-starting' of bottom-up designed synthetic cells can be circumvented (Figure 1.1). By introducing genes or synthetic cell modules, the functions of mitochondria can be expanded, or the autonomy of mitochondria can be increased, providing a system to test and build (parts of) a synthetic cell. *S. cerevisiae* is an optimal host for this approach, as it can sustain modifications and deletions in both the nuclear and the mitochondrial genomes [88, 89]. Before yeast mitochondria can be converted into synthetic cells, there are knowledge and technological gaps that require more research to be filled. Firstly, mitochondrial gene expression is heavily dependent on post-transcriptional regulation, which is not yet fully understood [83, 90]. Secondly, mitochondria are poorly genetically accessible, so efficient DNA editing technologies for mitochondria must be developed before the organelle can be used for synthetic cell research.

1.2 Mitochondrial biology of *Saccharomyces cerevisiae*

Mitochondrial functions and physiology

Mitochondria are double-membraned organelles present in almost all eukaryotes [91, 92], and fulfill many functions that are essential for the cell. The inside of mitochondria, the mitochondrial matrix, is enclosed by a protective outer membrane (OM) and a heavily folded inner membrane (IM) coated with proteins. These folded structures, known as cristae, provide a large active surface area within the relatively small mitochondrial volume (Figure 1.2) [93]. Mitochondria are popularly known as the 'powerhouse of the cell', as they are responsible for the conservation of most of the cellular metabolic energy in the form of ATP through oxidative phosphorylation using the electron-transport chain (ETC, or respiratory chain), located on the mitochondrial inner membrane. Mitochondria also conserve energy through the TCA cycle, which is located in the mitochondrial matrix [94, 95]. In addition, mitochondria fulfil many more essential functions in the cell, including the metabolism of reactive oxygen species (ROS), biosynthesis of fatty acids, amino acids, and co-factors, such as iron-sulfur clusters and heme, as well as the maintenance of redox homeostasis in the cell [96-99]. Additionally, mitochondria play a key role in cellular processes, primarily chronological aging, and apoptosis [100, 101]. Mitochondria are dynamic organelles and their morphology and mass shift

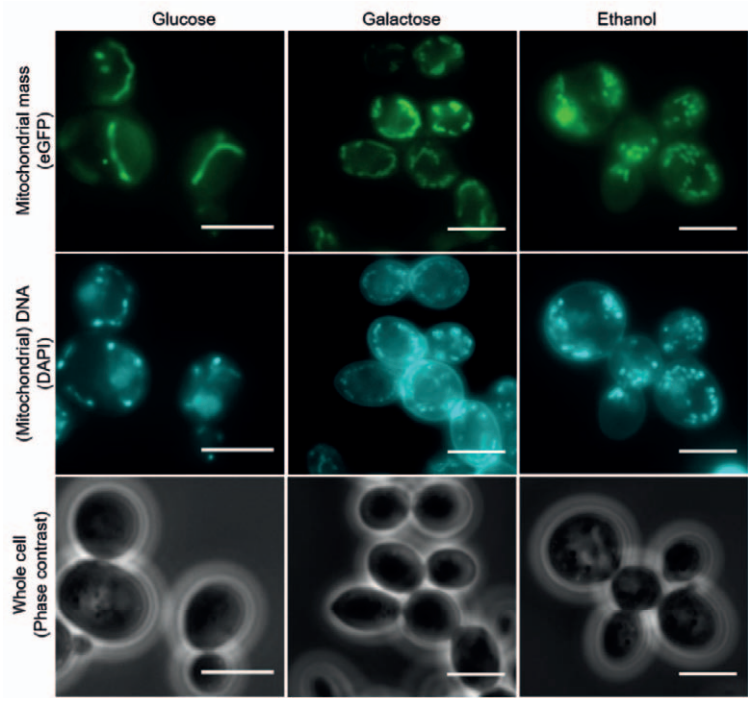


Figure 1.3. Mitochondrial morphology and DNA content of *S. cerevisiae* under different conditions. Mitochondrial mass is shown in green by targeting of a green fluorescent protein to the mitochondrial matrix (top), while cellular DNA content is stained with the blue fluorescent compound DAPI (middle). The mitochondrial morphology and DNA content changes due to carbon repression and altered respiratory capacity, which depends on the carbon source. Mitochondrial morphology is governed by the type of carbon source. Respiro-fermentative growth is observed on the carbon sources glucose and galactose, and growth is fully respiratory on ethanol. Mitochondrial functions are repressed by carbon catabolite repression on glucose, yielding tubular mitochondria. The scale bars represent 5 µm. Photo credit: Kavish Kohabir.

between growth conditions; on respiratory media like ethanol, where the cell relies on ATP generation through the mitochondrial ETC, mitochondria are spherical and occupy up to 35 % of the cellular mass, while on respiro-fermentative media such as glucose mitochondria are tubular and occupy a maximum of 5 % of the cellular volume (Figure 1.3) [102, 103].

The *S. cerevisiae* mitochondrial genome and proteome

While mitochondria contain a small, well characterized genome that encodes only eight proteins, their proteome is anything but simple, as it consists of approximately 1000 – 1200 proteins, which is circa 15 - 20 % of the total *S. cerevisiae* proteome [105]. Approximately 30 % is directly involved in mitochondrial functions and maintenance, while the rest fulfils a biosynthetic role in the cell [103]. 99 % of the mitochondrial proteins are encoded on the nuclear genome and imported from the cytosol [96, 106]. The non-cytosolic 1 % of the mitochondrial proteome is encoded on the mitochondrial genome and translated in the mitochondrial matrix. The mitochondrial genome (mtDNA) of *S. cerevisiae* was fully characterized in 1998 by Foury and colleagues and consists of an 86.1 kilobase pairs (kb) long sequence that primarily encodes eight components of the respiratory chain and two ribosomal subunits [84]. It also contains a full set of mitochondrial tRNAs (Chapter 3 of this thesis), as the genetic code of mitochondria is distinct from the yeast cytosolic genetic code [84, 107].

Mitochondrial transcription in *S. cerevisiae*

S. cerevisiae mitochondria utilize a transcription machinery distinct from the cytosolic machinery. Conversely to the nuclear genes, mitochondrial genes are transcribed as polycistronic primary transcripts. Their expression is regulated by a nuclear-encoded RNA polymerase that binds a nonanucleotide promoter sequence [108], and the produced transcripts are not polyadenylated. Just like the mitochondrial mass and DNA content, RNA transcription of the mitochondrial genome is also heavily dependent on the carbon source, as the nuclear encoded proteins controlling mitochondrial transcription are subjected to glucose repression [109-111]. To regulate expression levels of the different genes of a single transcript, mitochondrial polycistronic transcripts are subjected to posttranscriptional processing. This occurs at conserved dodecamer sequences or by autocatalytic cleavage. These RNA processing events free the transcripts destined for translation and modulate transcript level and gene expression (Figure 1.4) [83, 112-114]. Besides RNA processing of the primary transcripts, several mitochondrial open reading frames (ORFs) additionally contain introns. Therefore, the RNA is also subjected to splicing to remove intron sequences and mature the exon mRNA. Three of these ORFs (*COX1*, *COB* and 21S rRNA) have a mosaic structure. In these genes, the introns also contain coding ORFs. These ORFs encode proteins that are involved in the splicing of their respective primary transcript (Figure 1.4) [115, 116]. Mitochondrial gene expression is highly dependent on splicing of mitochondrial mRNAs. There is a lot of speculation as to why the mitochondrial introns have not been lost in evolution of the *S. cerevisiae* genome, as introns appear to be dispensable for the host and the distribution of mitochondrial introns widely vary between the *Saccharomycetaceae* [85, 116, 117]. *S. cerevisiae* mtDNA introns can be classified in group I and group II introns, which are distinct from each other in terms of sequence conservation, secondary structure configuration, and helix shape [118]. Group I and II introns are both ribozymes capable of self-splicing *in vitro* but require accessory proteins for splicing *in vivo*. Both intron classes can act as mobile genetic elements that proliferate through a mechanism termed intron homing, which is the insertion of an intron in a previously intron-less gene after a genetic cross [115, 117, 119]. Group I intron splicing is catalysed via a series of guanosine-initiated transesterification reactions, leading to splicing of a linear intron, while group II intron splicing proceeds through a two-step transesterification mechanism, yielding a circular intron lariat molecule [115]. Although the mechanisms of intron splicing are well-characterized, the underlying reason of their presence in the mitochondrial genome is still unknown. Disruption of the splicing homeostasis can heavily alter mitochondrial physiology [120, 121], especially under non-respiratory conditions, suggesting that introns play a yet unknown but important role in regulating mitochondrial gene expression.

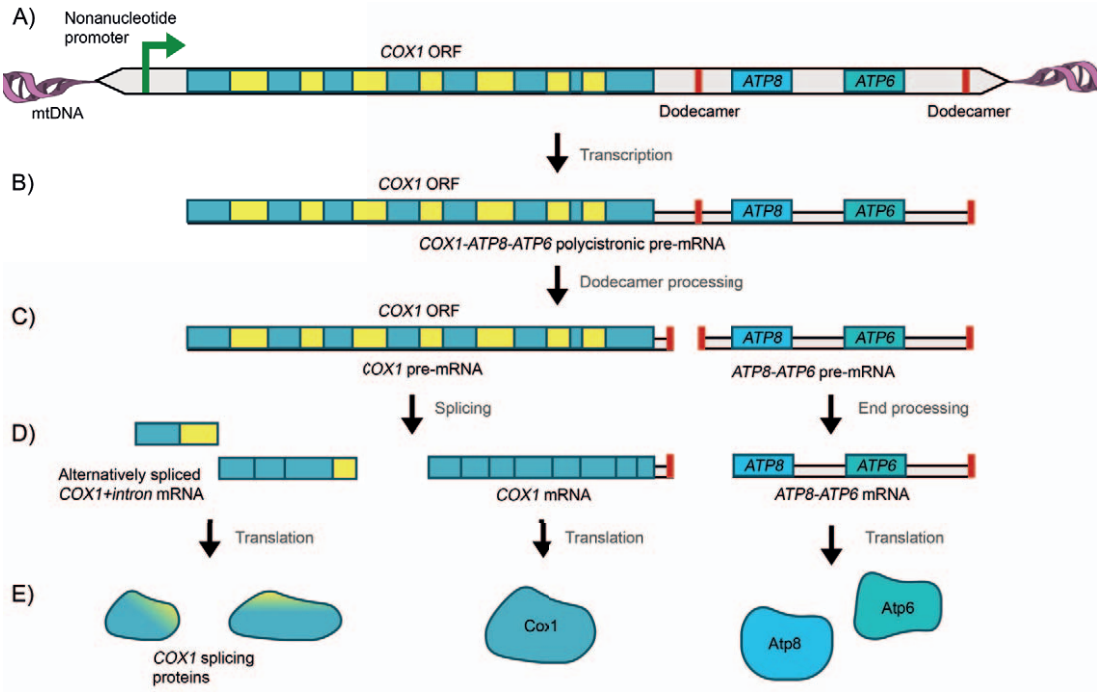


Figure 1.4. Different levels of polycistronic RNA processing present in the mitochondria, exemplified by the *COX1-ATP8-ATP6* polycistronic transcript. A) Mitochondrial gene expression is initiated by a nuclear-encoded RNA polymerase that binds a nonanucleotide promoter and is terminated at one of the dodecamer sequences. Mitochondrial genes can be a single open reading frame (blue, *ATP8/6*) or contain introns (yellow boxes, *COX1*). B) Multiple genes are expressed by a single promoter, yielding a polycistronic transcript. C) The transcript is further cleaved at internal dodecamer sites, yielding pre-mRNA, which still contains introns and sites for transcriptional regulation, such as (partial) dodecamers. D) The pre-mRNA is matured, either by splicing to remove any intron sequences or by end processing to remove any 5'-dodecamer sequences. (D) Intron-containing RNA can be spliced in different ways to yield different mature RNA transcripts, although most transcripts will be an exon-RNA devoid of introns. Mature mRNA can encode one or, in the case of *ATP8-6*, two proteins. E) Mature mRNAs are translated into their respective proteins.

1.3 Engineering the mitochondrial genome

Genome maintenance and engineering of *Saccharomyces cerevisiae*

The mitochondrial genome is packaged in nucleoids and is present in 50 to 200 copies, depending on the cellular environment, which represents up to 15 % of the whole cell DNA content (Figure 1.3) [122, 123]. The mtDNA is comprised of concatenated linear and circularized molecules, and despite the presence of three active origin of replications (*ori*s) on the mtDNA, mitochondria are hypothesized to replicate via an *ori*-independent rolling-circle mechanism, although this is still poorly understood [85, 124-126]. The highly dynamic organelles continually fuse and fragment, enabling mtDNA exchange between mitochondria within a cell. Oxidative phosphorylation releases ROS species that cause damage to mitochondria, and more particularly to mtDNA. The ROS-induced damage and fusing of mitochondria both contribute to heteroplasmy, which is the presence of multiple co-existing variants of mitochondrial genomes. The balance between mitochondrial division and fusion is, together with the rolling-circle amplification, thought to enable inheritance of fittest mitochondria by the daughter cell during cell division and to counteract the occurrence of respiratory deficiencies resulting from heteroplasmy [127-129]. Additionally, yeast and mammalian mitochondria can repair damaged mtDNA through homologous recombination, likely by resection and annealing using proteins from the *RAD52* epistasis group that are also responsible for the efficient homologous recombination of the yeast nuclear DNA [130-132].

Traditionally, mitochondria have been considered genetically inaccessible: while transfection or transformation methods to localize exogenous DNA in the nucleus are available, the DNA delivery in the mitochondria is difficult to achieve as they are not naturally competent for DNA uptake *in vivo*. Therefore, studies have traditionally relied on random chemical mutagenesis [133], and on generation of transmitochondrial cytoplasmic hybrid cells (cybrids) by fusion of recipient cells devoid of mtDNA with donor cells harboring mtDNA of interest [134, 135]. Most mitochondrial engineering efforts have focused on mammalian mitochondria, as mutated mtDNA is the source of many diseases for which no therapy exists [136]. However, mammalian cell lines cannot survive extensive mutations to their mtDNA, as the mtDNA-encoded respiratory chain is essential for energy conservation in these cells [137]. The ability of *S. cerevisiae* to live without a functional respiratory chain is a strong attribute for mtDNA engineering. *S. cerevisiae* is a Crabtree-positive yeast that can generate ATP either through respiration using the mitochondrially-expressed and localized respiratory chain, or through cytosolic substrate-level phosphorylation (i.e. ethanolic fermentation) [138]. Unlike mammalian cells, yeast can therefore survive damage to its mtDNA ranging from mutations or partial deletions (the 'petite' phenotype, or ρ^-) to complete loss of the mtDNA (ρ^0 phenotype) [87, 137]. The robustness of *S. cerevisiae* to mitochondrial mutations and its tolerance to harsh transformation protocols makes it one of the two organisms, together with the alga *Chlamydomonas reinhardtii*, in which mitochondrial DNA editing beyond the introduction of single base pair mutations is possible [88, 139].

Traditional mtDNA engineering approaches of yeast

S. cerevisiae mitochondria have been successfully engineered *in vivo* with the use of microprojectile bombardment or 'biolistic' transformation, in which ρ^0 or ρ^- cells are bombarded with microscopic metal particles coated with DNA, bypassing the lack of natural competence of mitochondria. In a small fraction of the cells, the DNA-coated particles penetrate not only the cell but also the mitochondria (Figure 1.5). The DNA is maintained in the mitochondria as a plasmid and mated with another mitochondrial mutant to complement a mitochondrial DNA defect, or directly integrated in the mitochondrial genome by homology-mediated repair [88, 139, 140]. Although this method can be used to efficiently engineer yeast mitochondria, it can only be used to integrate a limited number of genes, as a genetic selection pressure for the engineered mitochondria is required. This can be achieved in two ways. Firstly, mitochondria with a defined deletion in any of the respiratory chain genes can be used to integrate new genes. The respiratory defect can be complemented by integration of a DNA fragment encoding the deleted gene, as well as any other exogenous genes that are to be integrated, such as a fluorescent protein [141, 142]. However, this strategy can only be used once to integrate a gene, as subsequent gene integrations require additional defined deletions in the mitochondrial genome that can only be obtained by another round of biolistic transformation. The second strategy does not require prior modification of the mtDNA and relies on the use of selectable marker genes. Currently, two of such marker genes exist. The first is a dominant marker 'Barstar', in which an aspecific mitochondrial RNase is counteracted by the successful expression of Barstar from the mtDNA [143]. Secondly, mitochondria host many of the amino acid biosynthetic pathways, leveraged by the second marker system, which relies on auxotrophy-based screening. Currently, the only mitochondrial auxotrophic marker available is *ARG8*, encoding acetylornithine aminotransferase, which catalyses the essential step from *N*-acetyl ornithine to L-ornithine in the mitochondrially localized arginine biosynthetic pathway [144]. Deletion of the nuclear allele of *ARG8* can be complemented by integration of the mitochondrially recoded gene in the mitochondrial genome ([145], Chapter 4 of this thesis). The availability of the Barstar and *ARG8* markers expand the possibilities of mitochondrial genome engineering slightly, but currently the integration of multiple genes still requires continuous recycling of markers.

The CRISPR-Cas9 DNA-editing revolution

Limitations in editing the mitochondrial genome, such as the requirement for harsh transformation methods, ρ^- strains and the lack of mitochondrial markers, could be overcome by using clustered regularly interspaced short palindromic repeats (CRISPR) combined with RNA-guided endonucleases (RGENs) such as Cas9. CRISPR/RGEN is a part of the native immune system of many bacterial species against bacteriophages, that target and cut specific pathogenic DNA or RNA sequences [146, 147]. The realization that the target of Cas9 could be programmed by altering the guide RNA (gRNA) sequences in the CRISPRs [148, 149] has turned CRISPR/RGEN-based systems into revolutionary and indispensable genome editing tools. The

development of CRISPR/RGEN-based DNA editing systems allows fast and precise engineering of many different hosts, including previously poorly accessible genomes [150, 151]. CRISPR/RGEN editing relies on the expression of a gRNA component with a sequence complementary to the targeted DNA locus, and an RGEN, such as Cas9. The gRNA, bound by the RGEN, anneals to the DNA sequence complementary to the gRNA, upon which the RGEN introduces a double-strand break (DSB) in the genome at the intended location [148]. The DSB can then be repaired by the repair machinery of the host cell, such as non-homologous end-joining (NHEJ), which often results in mutation or deletion of one or a few base pairs, or homology-mediated repair (HR), where a DNA sequence with homology upstream and downstream to the DSB is integrated in the genome [152]. As mitochondria express an HR-repair mechanism, introduction of a DSB using a CRISPR/RGEN system and HR-mediated repair would be a feasible method of introducing new genes in the mitochondrial genome. Targeting a CRISPR/RGEN system to the mitochondria would require the localization of three components to the mitochondrial matrix: (i) an RNA-programmable endonuclease protein, (ii) a gRNA molecule to guide the endonuclease to the intended locus on the mtDNA and (iii) a DNA repair fragment, with a sequence that will define the type of DNA modification (e.g., base-editing, deletion or integration).

Targeted mitochondrial endonuclease proteins

Since 99 % of the mitochondrial proteins are imported, mitochondria harbor an efficient protein import machinery, consisting of the translocase of the outer mitochondrial membrane (TOM) and the translocase of the inner membrane (TIM). The TOM complex recognizes mitochondrial precursor proteins and unfolds and imports them into the mitochondrial intramembrane space (IMS), upon which the TIM complex directs the protein precursor towards either its insertion in the inner mitochondrial membrane (IMM) or the mitochondrial matrix (Figure 1.5). Mitochondrial protein precursors all contain an N-terminal positively charged amphipathic α-helix, called a presequence or mitochondrial targeting sequence (MTS), which is recognized by the TOM complex. The type of MTS determines the final sub-mitochondrial localization of a protein and is cleaved off upon import [153-156]. Programmable nucleases can therefore be targeted to the mitochondrial matrix by the addition of a specific N-terminal MTS. Many of the mitochondrial pathologies are caused by single base pair mutations or heteroplasmy of the mitochondrial DNA [157]. Therefore, many studies have focused on base-editing or depletion of erroneous mitochondrial DNA in a pathological context. Because of this, several mitochondrial endonucleases and base-editors that do not require RNA guides or DNA repairs have been developed. This includes mitochondrially targeted restriction endonucleases (mitoRE [158, 159]), TAL-Effector nucleases (mitoTALENs, [160]), zinc-finger nucleases (mitoZFNs [161, 162]) and homing endonucleases, as well as a mitochondrial cytosine base-editor (mitoDdCBE, [163]) and TALE-linked deaminases (TALED, [164], further reviewed in [165, 166]). Similarly, an RGEN can be targeted to the mitochondrial matrix through addition of an MTS. However, the integration of genes in the mtDNA using an RGEN requires not only a DSB in the mtDNA, but also the repair of the mtDNA with exogenous DNA, and import of a gRNA to guide the RGEN to the locus being edited.

DNA targeting to the mitochondria

Although isolated mitochondria appear competent for DNA uptake [167], no *in vivo* import mechanisms have presently been identified, complicating the integration of exogenous DNA in mitochondria (Figure 1.5) [168]. Although DNA can be delivered by means of the earlier described biolistic transformation, a less aggressive method compatible with standard transformation or transfection protocols would be desired. Despite the lack of natural competence of mitochondria, such methods for DNA delivery to the mitochondrial matrix have been developed. MTS peptide-DNA conjugates were shown to be targeted to the mitochondria, as well as lipophilic carriers, nanoparticles or dequalinium-based vesicles containing DNA, all of which have successfully been targeted to the mitochondrial matrix [169-171]. These however, are not compatible with genome engineering due to various reasons, including toxicity, hydrolysis of the conjugated DNA or incompatibility with *in vivo* DNA engineering [168, 172-174].

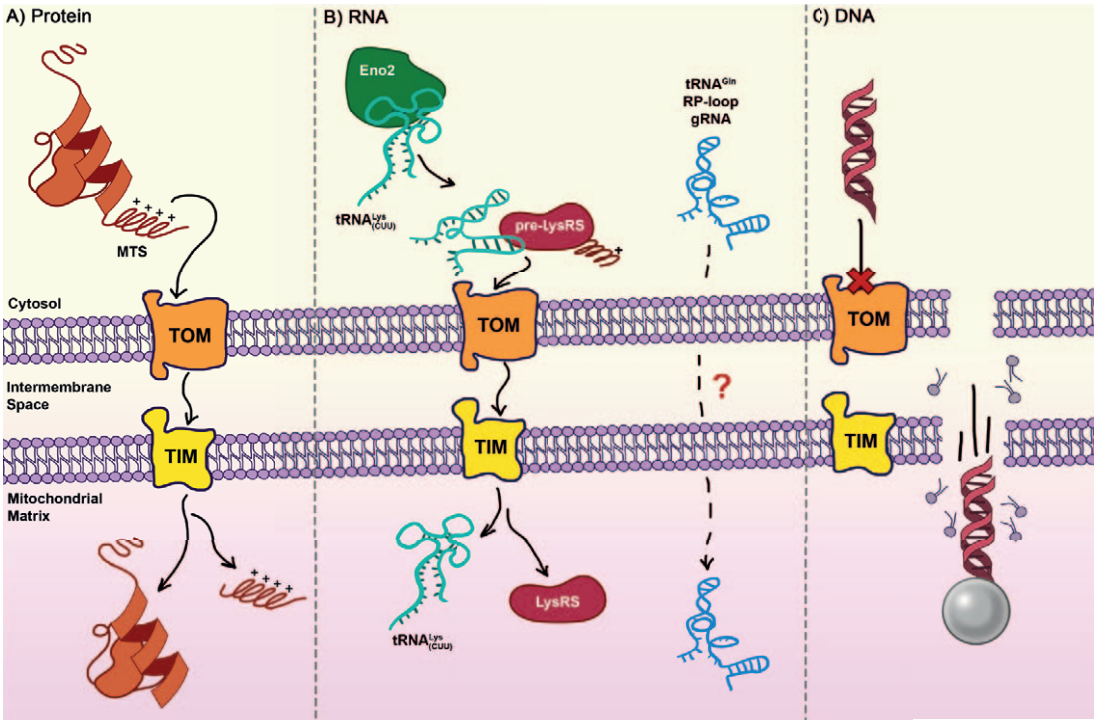


Figure 1.5. Targeting of proteins, RNA and DNA in mitochondria of *Saccharomyces cerevisiae*. A) Proteins that carry an N-terminal positively charged amphipathic α-helix (mitochondrial targeting sequence, MTS) are recognized and imported by the translocase of the outer membrane (TOM) and translocase of the inner membrane (TIM), which also cleaves the MTS. B) Proposed RNA import of tRNA^{Lys}(CUU) in yeast mitochondria. The tRNA is recruited to the mitochondria by Eno2, where it binds the precursor for mitochondrial lysyl-tRNA synthetase (pre-LysRS), undergoes a conformational change and is co-imported with the protein. Other tRNAs, RNase P (RP) RNA and gRNA were also reported to be imported by mitochondria, but the mechanism is unknown. C) DNA import in yeast mitochondria has not been reported *in vivo*, but DNA can be delivered in the mitochondrial matrix by high-velocity microprojectile bombardment with DNA-coated particles.

Targeting RNA to the mitochondria

To function in the context of genome editing, a mitochondrial RGEN would require the presence of gRNA in the mitochondrial matrix. Native RNA import has been described in mitochondria in multiple instances. In yeast, lysine tRNA^{Lys}(CUU) and glutamine tRNA^{Gln}(CUG) and tRNA^{Gln}(UUG) have been described to be imported *in vivo* in mitochondria [175, 176]. In mammals tRNA^{Gln}, the ribonucleic subunits of RNase P and RNase MRP, and 5S rRNA are imported in the mitochondria [177-180]. Mitochondria of several fungi and plants require import of the majority of mitochondrial tRNA, and *Trypanosomatidae* even import all mitochondrial tRNAs from the cytosol [181, 182]. Despite the widespread occurrence of mitochondrial (t)RNA import, exact mechanisms are yet to be unraveled. For *S. cerevisiae* tRNA^{Lys}(CUU), it is known that the import of the tRNA depends on sequential interactions with the glycolytic protein Eno2 and the mitochondrial lysyl-tRNA synthetase (pre-LysRS). The tRNA first binds Eno2p, which is recruited to glycolytic complexes on the mitochondrial membrane, where the tRNA associates with pre-LysRS which is translated in ribosomes located on the outside of the outer mitochondrial membrane. The tRNA then likely undergoes a conformational change and is co-imported into the mitochondria with pre-LysRS (Figure 1.5) [183-185]. This import mechanism was only observed for tRNA^{Lys}(CUU) and not for any of its cognate tRNAs, implying the tRNA and pre-LysRS have an interaction that is specific to the sequence or structure of the tRNA [186]. The import mechanisms of other (t)RNA species have not been elucidated yet, although proteins have been identified that reportedly assist in RNA import, including human RNA degradation protein PNPase, which was described to facilitate mRNA-sized RNA import *in vivo* and *in vitro* in human mitochondria [187, 188], while human rhodanese promotes 5S-RNA import in human mitochondria [189].

As discussed above, DNA is not readily imported in the mitochondrial matrix. Nevertheless, DNA-editing strategies exist that rely on reverse transcription of RNA as a repair template for double-stranded breaks [190-192]. Instead of supplying DNA itself for DSB repair, the repair DNA fragment could be generated by targeting an RNA-template to the mitochondrial matrix where it would be reverse transcribed into repair DNA. With such a system, CRISPR/RGEN editing in mitochondria could be achieved by targeting two proteins, Cas9 and a reverse transcriptase, to the mitochondrial matrix along with a chimeric gRNA/repair RNA molecule (Figure 1.6). Considering this approach, the success of using CRISPR/RGEN to engineer the mitochondrial DNA would strongly rely on the ability to target RNA into the mitochondrial matrix.

CRISPR/Cas9 editing of mtDNA: the *status quo*

Since the import of (t)RNA in mitochondria appears to rely on specific protein-RNA interactions, a variety of RNA-import signals have been identified and used to improve mitochondrial RNA import, mostly based on the sequences of yeast $tRNA^{Lys}$(CUU) and mammalian 5S rRNA and RNase (M)RP. Although the import of these sequences was shown to improve RNA import in isolated mitochondria *in vitro* [187, 188, 193-197], the *in vivo* efficiency remains controversial. Regardless, these signals have been used to target gRNA sequences into the mitochondrial matrix in an effort to edit the mtDNA using Cas9. A number of these studies reported depletion of mitochondrial DNA copies when repair DNA was not supplied, regardless of whether the gRNA carried mitochondrial RNA import signal or not [195, 198-201]. Other studies reported CRISPR/Cas9-mediated knock-ins, albeit with very low efficiency [202, 203]. However, one study demonstrated that the same single DNA modification was found when targeting Cas9 to the mitochondrial matrix regardless of the used gRNA [204]. This observation, combined with the lack of conclusive evidence and reproducibility on previous mitochondrial CRISPR/Cas9 studies [165, 205, 206], shows that efficient and stable mtDNA engineering with Cas9 has not been established yet, and that more research is required into RNA import, DNA repair and Cas9 activity in the mitochondria.

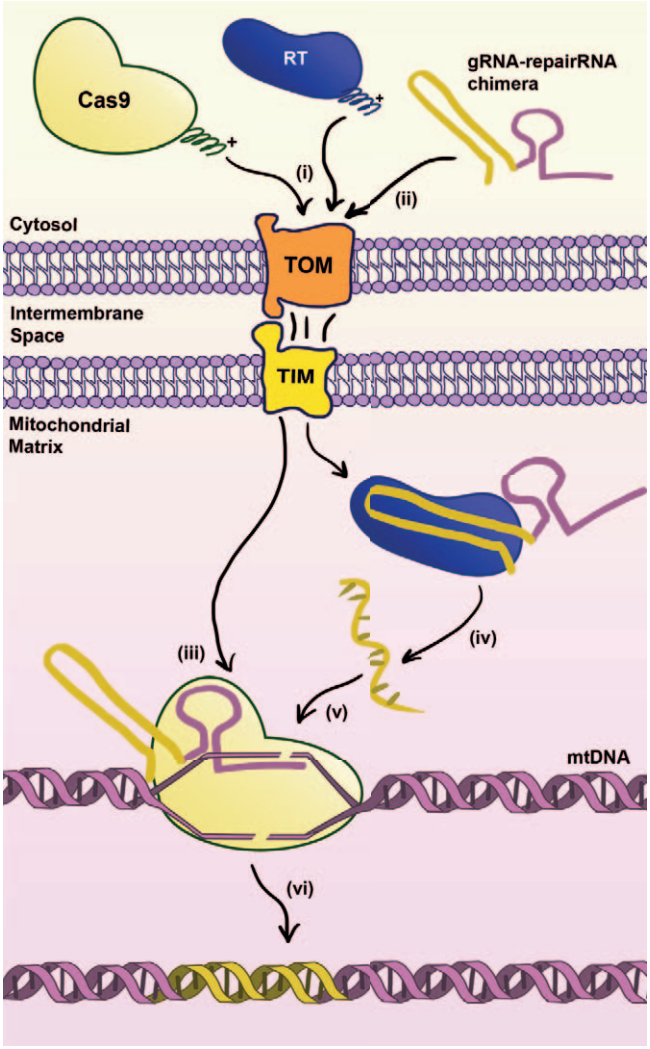


Figure 1.6. Proposed strategy for Cas9-mediated gene integration in the mitochondrial DNA while omitting a DNA-repair. Cas9 and a reverse transcriptase (RT) are targeted to the mitochondrial matrix using a mitochondrial targeting signal (i), while a chimeric RNA molecule consisting of a gRNA and an RNA-template for repair DNA is targeted through an RNA-uptake mechanism (ii). The Cas9-gRNA complex anneals to the targeted DNA-sequence in the mitochondrial DNA (mtDNA), where it induces a double-stranded break (DSB, iii). Simultaneously, a copy of the chimeric RNA is used by the RT to generate a single stranded DNA (ssDNA) repair template (iv), which is used for homology-mediated repair of the DSB (v), resulting in stable integration of the ssDNA in the mtDNA (vi).

1.4 Scope of this thesis

With the rise of synthetic biology, editing genomes in a modular way has become a staple in biotechnology. The next step goes beyond engineering living organisms, but aims at building entire cells *de novo*, by combining basic macromolecules into synthetic cells. At the core of any synthetic cell will be its genome, containing all the genes required for the division and maintenance of the cell, as well as any functions that the synthetic cells have.
The overarching goal of the presented research was to establish strategies for building the genome of a synthetic cell using *Saccharomyces cerevisiae*, by either building a genome *de novo* or by engineering and expanding the genome of mitochondria, as an intermediate between top-down and bottom-up biology.

Bottom-up construction of the genome of a synthetic cell requires the assembly of DNA containing transcriptional units encoding all functions of the synthetic cell. The assembly of DNA of such size could be achieved by utilizing the homology-based DNA repair system of *S. cerevisiae*. **Chapter 2** summarizes the contributions *S. cerevisiae* has made to efficient large-scale genome assembly, what the strengths and shortcomings of this technology are, and envisions what role yeast may play in the future as a 'genome foundry'.

At present, the construction of synthetic cells is not advanced to the level where an assembled genome can readily be encapsulated and expressed. The compatibility of yeast *in vivo* assembly machinery with exogenous and/or repetitive DNA sequences requires further assessment. On top of that, there are still many fundamental questions that need resolving, including the minimal set of genes required to obtain a dividing cell, and how to 'kick-start' the cell [69, 70]. To bypass these challenges, we envisioned the use of mitochondria as an intermediate to building synthetic cells. The semi-autonomous and well-characterized organelle contains its own membrane, genome, transcription and translation machinery. Its functions and autonomy could be engineered by expanding the mitochondrial genome.

Expanding the mitochondrial genome does not only require good understanding of the mitochondrial genome sequence, but also on its expression. Although the mitochondrial genome is relatively small, it is highly complex due to its polycistronic nature and is subject to splicing and co- and post-transcriptional processing. **Chapter 3** presents a method based on the isolation of mitochondria and Nanopore long-read sequencing of mitochondrial RNA to characterize the mitochondrial transcriptome and its processing.

The integration of genes in the mitochondrial genome does not only require characterization of the genome, but also identification of marker sequences that can be used to screen for successful integration in the mitochondrial genome. Currently *ARG8*, an essential gene in arginine biosynthesis, is the only auxotrophic marker available for selection of mitochondrial expression. **Chapter 4** reports the discovery of a frequently occurring mutation that can bypass arginine auxotrophy in a *Δarg8* background, and the development of a strategy to circumvent the loss of arginine auxotrophy.

Besides the availability of auxotrophic markers, mitochondrial genome editing is hampered by the lack of characterization of exogenous RNA import, which is required for CRISPR/RNA-guided endonuclease engineering. In **Chapter 5**, a systematic analysis of translocation of mRNA to the mitochondrial matrix was performed using a fluorescence-based method. Additionally, an evolutionary engineering approach, based on import of the *ARG8* marker characterized in **Chapter 4**, was applied in an attempt to improve RNA import, and shine a light on the elusive underlying mechanism.

References

1. Cameron, D.E., Bashor, C.J., and Collins, J.J., A brief history of synthetic biology. *Nature Reviews Microbiology*, *2014*. ***12****(5): p. 381-390*.
2. French, K.E., Harnessing synthetic biology for sustainable development. *Nature Sustainability*, *2019*. ***2****(4): p. 250-252*.
3. Flores Bueso, Y. and Tangney, M., Synthetic biology in the driving seat of the bioeconomy. *Trends in Biotechnology*, *2017*. ***35****(5): p. 373-378*.
4. El Karoui, M., Hoyos-Flight, M., and Fletcher, L., Future trends in synthetic biology—a report. *Frontiers in Bioengineering and Biotechnology*, *2019: p. 175*.
5. Hassan, M., Naz, A., Siddique, A., Shahzadi, S., Din, S.u., Yaseen, Z., Naqvi, S., Ali, Q., Awan, F.M., and Ikram, A., Synthetic biology in healthcare: technologies and applications, in Biotechnology in Healthcare, D. Barh, Editor. 2022, *Academic Press*. p. 41-53.
6. Tan, X., Letendre, J.H., Collins, J.J., and Wong, W.W., Synthetic biology in the clinic: engineering vaccines, diagnostics, and therapeutics. *Cell*, *2021*. ***184****(4): p. 881-898*.
7. Lienert, F., Lohmueller, J.J., Garg, A., and Silver, P.A., Synthetic biology in mammalian cells: next generation research tools and therapeutics. *Nature Reviews Molecular Cell Biology*, *2014*. ***15****(2): p. 95-107*.
8. Pedrolli, D.B., Ribeiro, N.V., Squizato, P.N., de Jesus, V.N., Cozetto, D.A., Tuma, R.B., Gracindo, A., Cesar, M.B., Freire, P.J.C., da Costa, A.F.M., Lins, M.R.C.R., Correa, G.G., and Cerri, M.O., Engineering microbial living therapeutics: The synthetic biology toolbox. *Trends in Biotechnology*, *2019*. ***37****(1): p. 100-115*.
9. Katz, L., Chen, Y.Y., Gonzalez, R., Peterson, T.C., Zhao, H., and Baltz, R.H., Synthetic biology advances and applications in the biotechnology industry: a perspective. *Journal of Industrial Microbiology and Biotechnology*, *2018*. ***45****(7): p. 449-461*.
10. Clarke, L. and Kitney, R., Developing synthetic biology for industrial biotechnology applications. *Biochemical Society Transactions*, *2020*. ***48****(1): p. 113-122*.
11. Augustin, M.A., Hartley, C.J., Maloney, G., and Tyndall, S., Innovation in precision fermentation for food ingredients. *Critical Reviews in Food Science and Nutrition*, *2023: p. 1-21*.
12. Dellomonaco, C., Fava, F., and Gonzalez, R., The path to next generation biofuels: successes and challenges in the era of synthetic biology. *Microbial cell factories*, *2010*. ***9****(1): p. 3*.
13. Jullesson, D., David, F., Pfleger, B., and Nielsen, J., Impact of synthetic biology and metabolic engineering on industrial production of fine chemicals. *Biotechnology Advances*, *2015*. ***33****(7): p. 1395-1402*.
14. Wurtzel, E.T., Vickers, C.E., Hanson, A.D., Millar, A.H., Cooper, M., Voss-Fels, K.P., Nikel, P.I., and Erb, T.J., Revolutionizing agriculture with synthetic biology. *Nature Plants*, *2019*. ***5****(12): p. 1207-1210*.
15. Roell, M.-S. and Zurbriggen, M.D., The impact of synthetic biology for future agriculture and nutrition. *Current opinion in biotechnology*, *2020*. ***61****: p. 102-109*.
16. Varshney, R.K., Bansal, K.C., Aggarwal, P.K., Datta, S.K., and Craufurd, P.Q., Agricultural biotechnology for crop improvement in a variable climate: hope or hype? *Trends in Plant Science*, *2011*. ***16****(7): p. 363-371*.
17. Sargent, D., Conaty, W.C., Tissue, D.T., and Sharwood, R.E., Synthetic biology and opportunities within agricultural crops. *Journal of Sustainable Agriculture and Environment*, *2022*. ***1****(2): p. 89-107*.
18. Zhang, G., Zhao, X., Li, X., Du, G., Zhou, J., and Chen, J., Challenges and possibilities for bio-manufacturing cultured meat. *Trends in Food Science & Technology*, *2020*. ***97****: p. 443-450*.
19. Guindani, C., da Silva, L.C., Cao, S., Ivanov, T., and Landfester, K., Synthetic cells: From simple bio-inspired modules to sophisticated integrated systems. *Angewandte Chemie International Edition*, *2022*. ***61****(16): p. e202110855*.
20. Stano, P., Is research on "synthetic cells" moving to the next level? *Life*, *2019*. ***9****(1): p. 3*.
21. Göpfrich, K., Platzman, I., and Spatz, J.P., Mastering Complexity: Towards bottom-up construction of multifunctional eukaryotic synthetic cells. *Trends in Biotechnology*, *2018*. ***36****(9): p. 938-951*.
22. Gaut, N.J. and Adamala, K.P., Reconstituting natural cell elements in synthetic cells. *Advanced Biology*, *2021*. ***5****(3): p. 2000188*.
23. Spoelstra, W.K., Deshpande, S., and Dekker, C., Tailoring the appearance: what will synthetic cells look like? *Current opinion in biotechnology*, *2018*. ***51****: p. 47-56*.
24. Olivi, L., Berger, M., Creyghton, R.N.P., De Franceschi, N., Dekker, C., Mulder, B.M., Claassens, N.J., ten Wolde, P.R., and van der Oost, J., Towards a synthetic cell cycle. *Nature communications*, *2021*. ***12****(1): p. 4531*.
25. Robinson, A.O., Venero, O.M., and Adamala, K.P., Toward synthetic life: Biomimetic synthetic cell communication. *Current Opinion in Chemical Biology*, *2021*. ***64****: p. 165-173*.
26. van Nies, P., Westerlaken, I., Blanken, D., Salas, M., Mencia, M., and Danelon, C., Self-replication of DNA by its encoded proteins in liposome-based synthetic cells. *Nature Communications*, *2018*. ***9****(1): p. 1583*.
27. Herianto, S., Chien, P.-J., Ho, J.-a.A., and Tu, H.-L., Liposome-based artificial cells: From gene expression to reconstitution of cellular functions and phenotypes. *Biomaterials Advances*, *2022*. ***142****: p. 213156*.
28. Blanken, D., Foschepoth, D., Serrão, A.C., and Danelon, C., Genetically controlled membrane synthesis in liposomes. *Nature communications*, *2020*. ***11****(1): p. 1-13*.
29. Bailoni, E. and Poolman, B., ATP recycling fuels sustainable glycerol 3-phosphate formation in synthetic cells fed by dynamic dialysis. *ACS synthetic biology*, *2022*. ***11****(7): p. 2348-2360*.
30. Partipilo, M., Ewins, E.J., Fralliccardi, J., Robinson, T., Poolman, B., and Slotboom, D.J., Minimal pathway for the regeneration of redox cofactors. *JACS Au*, *2021*. ***1****(12): p. 2280-2293*.
31. Kleineberg, C., Wölfer, C., Abbasnia, A., Pischel, D., Bednarz, C., Ivanov, I., Heitkamp, T., Börsch, M., Sundmacher, K., and Vidaković-Koch, T., Light-driven ATP regeneration in diblock/grafted hybrid vesicles. *ChemBioChem*, *2020*. ***21****(15): p. 2149-2160*.
32. Ghosh, S., Mohajerani, F., Son, S., Velegol, D., Butler, P.J., and Sen, A., Motility of enzyme-powered vesicles. *Nano letters*, *2019*. ***19****(9): p. 6019-6026*.
33. Chen, S., Sun, Z., and Murrell, M., *In vitro* reconstitution of the actin cytoskeleton inside giant unilamellar vesicles. *Journal of Visualized Experiments*, *2022*. ***186****: p. e64026*.
34. Lavickova, B., Laohakunakorn, N., and Maerkl, S.J., A partially self-regenerating synthetic cell. *Nature communications*, *2020*. ***11****(1): p. 6340*.
35. Has, C. and Sunthar, P., A comprehensive review on recent preparation techniques of liposomes. *Journal of Liposome Research*, *2020*. ***30****(4): p. 336-365*.
36. Powell, K., Biology from scratch. *Nature*, *2018*. ***563****: p. 172-175*.
37. Einfalt, T., Garni, M., Witzigmann, D., Sieber, S., Baltisberger, N., Huwyler, J., Meier, W., and Palivan, C.G., Bioinspired molecular factories with architecture and *in vivo* functionalities as cell mimics. *Advanced Science*, *2020*. ***7****(4): p. 1901923*.
38. Pohorille, A. and Deamer, D., Artificial cells: prospects for biotechnology. *Trends in Biotechnology*, *2002*. ***20****(3): p. 123-128*.
39. LeDuc, P.R., Wong, M.S., Ferreira, P.M., Groff, R.E., Haslinger, K., Koonce, M.P., Lee, W.Y., Love, J.C., McCammon, J.A., Monteiro-Riviere, N.A., Rotello, V.M., Rubloff, G.W., Westervelt, R., and Yoda, M., Towards an i*n vivo* biologically

inspired nanofactory. *Nature Nanotechnology*, *2007*. ***2****(1): p. 3-7*.

40. Chang, T.M.S., Future prospects for artificial blood. *Trends in Biotechnology*, *1999*. ***17****(2): p. 61-67*.
41. Krinsky, N., Kaduri, M., Zinger, A., Shainsky-Roitman, J., Goldfeder, M., Benhar, I., Hershkovitz, D., and Schroeder, A., Synthetic cells: Synthetic cells synthesize therapeutic proteins inside tumors *Advanced Healthcare Materials*, *2018*. ***7****(9): p. 1870038*.
42. Elani, Y., Trantidou, T., Wylie, D., Dekker, L., Polizzi, K., Law, R.V., and Ces, O., Constructing vesicle-based artificial cells with embedded living cells as organelle-like modules. *Scientific reports*, *2018*. ***8****(1): p. 4564*.
43. de Jong, B., Siewers, V., and Nielsen, J., Systems biology of yeast: enabling technology for development of cell factories for production of advanced biofuels. *Current opinion in biotechnology*, *2012*. ***23****(4): p. 624-630*.
44. Choe, D., Kim, S.C., Palsson, B.O., and Cho, B.-K., Construction of minimal genomes and synthetic cells, in Minimal cells: Design, construction, biotechnological applications, A.R. Lara and G. Gosset, Editors. 2020, *Springer International Publishing*: Cham. p. 45-67.
45. Völler, J.-S. and Budisa, N., Coupling genetic code expansion and metabolic engineering for synthetic cells. *Current opinion in biotechnology*, *2017*. ***48****: p. 1-7*.
46. Lartigue, C., Glass, J.I., Alperovich, N., Pieper, R., Parmar, P.P., Hutchison, C.A., Smith, H.O., and Venter, J.C., Genome transplantation in bacteria: Changing one species to another. *Science*, *2007*. ***317****(5838): p. 632-638*.
47. Xu, C., Hu, S., and Chen, X., Artificial cells: from basic science to applications. *Materials Today*, *2016*. ***19****(9): p. 516-532*.
48. Gibson, D.G., Glass, J.I., Lartigue, C., Noskov, V.N., Chuang, R.-Y., Algire, M.A., Benders, G.A., Montague, M.G., Ma, L., and Moodie, M.M., Creation of a bacterial cell controlled by a chemically synthesized genome. *Science*, *2010*. ***329****(5987): p. 52-56*.
49. Costello, A. and Badran, A.H., Synthetic biological circuits within an orthogonal central dogma. *Trends in Biotechnology*, *2021*. ***39****(1): p. 59-71*.
50. Budisa, N., Xenobiology, new-to-nature synthetic cells and genetic firewall. *Current Organic Chemistry*, *2014*. ***18****(8): p. 936-943*.
51. Neumann, H., Wang, K., Davis, L., Garcia-Alai, M., and Chin, J.W., Encoding multiple unnatural amino acids via evolution of a quadruplet-decoding ribosome. *Nature*, *2010*. ***464****(7287): p. 441-444*.
52. Wang, L., Jiang, S., Chen, C., He, W., Wu, X., Wang, F., Tong, T., Zou, X., Li, Z., Luo, J., Deng, Z., and Chen, S., Synthetic genomics: From DNA synthesis to genome design. *Angewandte Chemie International Edition*, *2018*. ***57****(7): p. 1748-1756*.
53. Meng, F. and Ellis, T., The second decade of synthetic biology: 2010–2020. *Nature communications*, *2020*. ***11****(1): p. 5174*.
54. Venter, J.C., Glass, J.I., Hutchison, C.A., and Vashee, S., Synthetic chromosomes, genomes, viruses, and cells. *Cell*, *2022*. ***185****(15): p. 2708-2724*.
55. Fredens, J., Wang, K., de la Torre, D., Funke, L.F., Robertson, W.E., Christova, Y., Chia, T., Schmied, W.H., Dunkelmann, D.L., and Beránek, V., Total synthesis of *Escherichia coli* with a recoded genome. *Nature*, *2019*. ***569****(7757): p. 514-518*.
56. Ostrov, N., Landon, M., Guell, M., Kuznetsov, G., Teramoto, J., Cervantes, N., Zhou, M., Singh, K., Napolitano, M.G., Moosburner, M.,... Church, G.M., Design, synthesis, and testing toward a 57-codon genome. *Science*, *2016*. ***353****(6301): p. 819-822*.
57. Venetz, J.E., Del Medico, L., Wölfle, A., Schächle, P., Bucher, Y., Appert, D., Tschan, F., Flores-Tinoco, C.E., van Kooten, M., and Guennoun, R., Chemical synthesis rewriting of a bacterial genome to achieve design flexibility and biological functionality. *Proceedings of the National Academy of Sciences*, *2019*. ***116****(16): p. 8070-8079*.
58. Eisenstein, M., How to build a genome. *Nature*, *2020*. ***578****(7796): p. 633-635*.
59. Gil, R., The minimal gene-set machinery, in Reviews in Cell Biology and Molecular Medicine. 2014. p. 1-36.
60. Dai, J., Boeke, J.D., Luo, Z., Jiang, S., and Cai, Y., Sc3.0: revamping and minimizing the yeast genome. *Genome Biology*, *2020*. ***21****(1)*.
61. Hutchison, C.A., Chuang, R.-Y., Noskov, V.N., Assad-Garcia, N., Deerinck, T.J., Ellisman, M.H., Gill, J., Kannan, K., Karas, B.J., and Ma, L., Design and synthesis of a minimal bacterial genome. *Science*, *2016*. ***351****(6280): p. aad6253*.
62. Larrimore, K.E. and Rancati, G., The conditional nature of gene essentiality. *Current Opinion in Genetics & Development*, *2019*. ***58-59****: p. 55-61*.
63. Juhas, M., Eberl, L., and Glass, J.I., Essence of life: essential genes of minimal genomes. *Trends in Cell biology*, *2011*. ***21****(10): p. 562-568*.
64. Schwille, P., Spatz, J., Landfester, K., Bodenschatz, E., Herminghaus, S., Sourjik, V., Erb, T.J., Bastiaens, P., Lipowsky, R., Hyman, A.,... Sundmacher, K., MaxSynBio: Avenues towards creating cells from the bottom up. *Angewandte Chemie International Edition*, *2018*. ***57****(41): p. 13382-13392*.
65. Ichihashi, N. and Stano, P., Trends and outlooks in synthetic biology: A special issue for celebrating 10 years of life and its landmarks. *Life*, *2022*. ***12****(2): p. 181*.
66. Ces, O. and Elani, Y., Community building in synthetic biology. *Experimental Biology and Medicine*, *2019*. ***244****(4): p. 281-282*.
67. Frischmon, C., Sorenson, C., Winikoff, M., and Adamala, K.P., Build-a-Cell: Engineering a synthetic cell community. *Life*, *2021*. ***11****(11): p. 1176*.
68. Mutschler, H., Robinson, T., Tang, T.Y.D., and Wegner, S., Special issue on bottom-Up synthetic biology. *ChemBioChem*, *2019*. ***20****(20): p. 2533-2534*.
69. Abil, Z. and Danelon, C., Roadmap to Building a Cell: An Evolutionary Approach. *Frontiers in Bioengineering and Biotechnology*, *2020*. ***8****: p. 927*.
70. Juhas, M., On the road to synthetic life: the minimal cell and genome-scale engineering. *Critical reviews in biotechnology*, *2016*. ***36****(3): p. 416-423*.
71. Forster, A.C. and Church, G.M., Towards synthesis of a minimal cell. *Molecular Systems Biology*, *2006*. ***2****: p. 45*.
72. Gil, R., Silva, F.J., Peretó, J., and Moya, A., Determination of the core of a minimal bacterial gene set. *Microbiology and Molecular Biology Reviews*, *2004*. ***68****(3): p. 518-537*.
73. Postma, E.D., Dashko, S., Lars, Shannara, K., Van Den Broek, M., Daran, J.-M., and Daran-Lapujade, P., A supernumerary designer chromosome for modular *in vivo* pathway assembly in *Saccharomyces cerevisiae*. *Nucleic acids research*, *2021*. ***49****(3): p. 1769-1783*.
74. Juhas, M. and Ajioka, J.W., High molecular weight DNA assembly *in vivo* for synthetic biology applications. *Crit Rev Biotechnol*, *2017*. ***37****(3): p. 277-286*.
75. Martin, W.F., Garg, S., and Zimorski, V., Endosymbiotic theories for eukaryote origin. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *2015*. ***370****(1678): p. 20140330*.
76. Gray, M.W., The endosymbiont hypothesis revisited, in International Review of Cytology, D.R. Wolstenholme and K.W. Jeon, Editors. 1992, *Academic Press*. p. 233-357.
77. Sabater, B., Evolution and function of the chloroplast. Current investigations and perspectives. *International Journal of Molecular Sciences*, *2018*. ***19****(10)*.
78. Archibald, John M., Endosymbiosis and eukaryotic cell evolution. *Current Biology*, *2015*. ***25****(19): p. R911-R921*.
79. Roger, A.J., Muñoz-Gómez, S.A., and Kamikawa, R., The origin and diversification of mitochondria. *Current Biology*, *2017*. ***27****(21): p. R1177-R1192*.
80. Gray, M.W., Burger, G., and Lang, B.F., The origin and early evolution of mitochondria. *Genome Biology*, *2001*. ***2****(6): p. reviews1018.1*.
81. Youle, R.J., Mitochondria - Striking a balance between host

and endosymbiont. *Science*, *2019*. ***365****(6454): p. eaaw9855*.

82. Taanman, J.-W., The mitochondrial genome: structure, transcription, translation and replication. *Biochimica et Biophysica Acta (BBA) - Bioenergetics*, *1999*. ***1410****(2): p. 103-123*.
83. Lipinski, K.A., Kaniak-Golik, A., and Golik, P., Maintenance and expression of the *S. cerevisiae* mitochondrial genome—From genetics to evolution and systems biology. *Biochimica et Biophysica Acta (BBA) - Bioenergetics*, *2010*. ***1797****(6): p. 1086-1098*.
84. Foury, F., Roganti, T., Lecrenier, N., and Purnelle, B., The complete sequence of the mitochondrial genome of *Saccharomyces cerevisiae*. *FEBS letters*, *1998*. ***440****(3): p. 325-331*.
85. Dujon, B., Mitochondrial genetics revisited. *Yeast*, *2020*. ***37****(2): p. 191-205*.
86. Shaw, J.M. and Nunnari, J., Mitochondrial dynamics and division in budding yeast. *Trends in Cell biology*, *2002*. ***12****(4): p. 178-184*.
87. Contamine, V. and Picard, M., Maintenance and integrity of the mitochondrial genome: A plethora of nuclear genes in the budding yeast. *Microbiology and Molecular Biology Reviews*, *2000*. ***64****(2): p. 281-315*.
88. Bonnefoy, N. and Fox, T.D., Directed alteration of *Saccharomyces cerevisiae* mitochondrial DNA by biolistic transformation and homologous recombination, in Mitochondria: Practical Protocols, D. Leister and J.M. Herrmann, Editors. 2007, *Humana Press*: Totowa, NJ. p. 153-166.
89. Mans, R., van Rossum, H.M., Wijsman, M., Backx, A., Kuijpers, N.G., van den Broek, M., Daran-Lapujade, P., Pronk, J.T., van Maris, A.J., and Daran, J.-M.G., CRISPR/Cas9: a molecular Swiss army knife for simultaneous introduction of multiple genetic modifications in *Saccharomyces cerevisiae*. *FEMS yeast research*, *2015*. ***15****(2)*.
90. Deshpande, I., Seeber, A., Shimada, K., Keusch, J.J., Gut, H., and Gasser, S.M., Structural basis of Mec1-Ddc2-RPA assembly and activation on single-stranded DNA at sites of damage. *Molecular cell*, *2017*. ***68****(2): p. 431-445. e5*.
91. Karnkowska, A., Vacek, V., Zubáčová, Z., Treitli, S.C., Petrželková, R., Eme, L., Novák, L., Žárský, V., Barlow, L.D., Herman, E.K.,... Hampl, V., A eukaryote without a mitochondrial organelle. *Current Biology*, *2016*. ***26****(10): p. 1274-1284*.
92. Hackstein, J.H.P., Baker, S.E., van Hellemond, J.J., and Tielens, A.G.M., Hydrogenosomes of anaerobic fungi: An alternative way to adapt to anaerobic environments, in Hydrogenosomes and Mitosomes: Mitochondria of Anaerobic Eukaryotes, J. Tachezy, Editor. 2019, *Springer International Publishing*: Cham. p. 159-175.
93. Zick, M., Rabl, R., and Reichert, A.S., Cristae formation—linking ultrastructure and function of mitochondria. *Biochimica et Biophysica Acta (BBA) - Molecular Cell Research*, *2009*. ***1793****(1): p. 5-19*.
94. Siekevitz, P., Powerhouse of the Cell. *Scientific American*, *1957*. ***197****(1): p. 131-144*.
95. Korla, K. and Mitra, C.K., Modelling the Krebs cycle and oxidative phosphorylation. *Journal of Biomolecular Structure and Dynamics*, *2014*. ***32****(2): p. 242-256*.
96. Malina, C., Larsson, C., and Nielsen, J., Yeast mitochondria: an overview of mitochondrial biology and the potential of mitochondrial systems biology. *FEMS yeast research*, *2018*. ***18****(5)*.
97. Lill, R. and Freibert, S.-A., Mechanisms of mitochondrial iron-sulfur protein biogenesis. *Annual review of biochemistry*, *2020*. ***89****(1): p. 471-499*.
98. Swenson, S.A., Moore, C.M., Marcero, J.R., Medlock, A.E., Reddi, A.R., and Khalimonchuk, O., From synthesis to utilization: The ins and outs of mitochondrial heme. *Cells*, *2020*. ***9****(3): p. 579*.
99. Kastaniotis, A.J., Autio, K.J., Kerätär, J.M., Monteuuis, G., Mäkelä, A.M., Nair, R.R., Pietikäinen, L.P., Shvetsova, A., Chen, Z., and Hiltunen, J.K., Mitochondrial fatty acid synthesis, fatty acids and mitochondrial physiology. *Biochimica et Biophysica Acta (BBA) - Molecular and Cell Biology of Lipids*, *2017*. ***1862****(1): p. 39-48*.
100. Pereira, C., Silva, R.D., Saraiva, L., Johansson, B., Sousa, M.J., and Côrte-Real, M., Mitochondria-dependent apoptosis in yeast. *Biochimica et Biophysica Acta (BBA) - Molecular Cell Research*, *2008*. ***1783****(7): p. 1286-1302*.
101. Jazwinski, S.M., Yeast longevity and aging—the mitochondrial connection. *Mechanisms of Ageing and Development*, *2005*. ***126****(2): p. 243-248*.
102. Visser, W., van Spronsen, E.A., Nanninga, N., Pronk, J.T., Kuenen, J.G., and van Dijken, J.P., Effects of growth conditions on mitochondrial morphology in *Saccharomyces cerevisiae*. *Antonie van Leeuwenhoek*, *1995*. ***67****(3): p. 243-253*.
103. Di Bartolomeo, F., Malina, C., Campbell, K., Mormino, M., Fuchs, J., Vorontsov, E., Gustafsson, C.M., and Nielsen, J., Absolute yeast mitochondrial proteome quantification reveals trade-off between biosynthesis and energy generation during diauxic shift. *Proceedings of the National Academy of Sciences*, *2020*. ***117****(13): p. 7524-7535*.
104. Lai, N., M. Kummitha, C., Rosca, M.G., Fujioka, H., Tandler, B., and Hoppel, C.L., Isolation of mitochondrial subpopulations from skeletal muscle: Optimizing recovery and preserving integrity. *Acta Physiologica*, *2019*. ***225****(2): p. e13182*.
105. Ho, B., Baryshnikova, A., and Brown, G.W., Unification of protein abundance datasets yields a quantitative *Saccharomyces cerevisiae* proteome. *Cell Systems*, *2018*. ***6****(2): p. 192-205.e3*.
106. Vögtle, F.N., Burkhart, J.M., Gonczarowska-Jorge, H., Kücükköse, C., Taskin, A.A., Kopczynski, D., Ahrends, R., Mossmann, D., Sickmann, A., Zahedi, R.P., and Meisinger, C., Landscape of submitochondrial protein distribution. *Nature communications*, *2017*. ***8****(1): p. 290*.
107. Fox, T.D., Natural variation in the genetic code. *Annual review of genetics*, *1987*. ***21****(1): p. 67-91*.
108. Biswas, T., Ticho, B., and Getz, G., *In vitro* characterization of the yeast mitochondrial promoter using single-base substitution mutants. *Journal of Biological Chemistry*, *1987*. ***262****(28): p. 13690-13696*.
109. Ulery, T.L., Jang, S.H., and Jaehning, J.A., Glucose repression of yeast mitochondrial transcription: kinetics of derepression and role of nuclear genes. *Molecular and Cellular Biology*, *1994*. ***14****(2): p. 1160-70*.
110. Mueller, D.M. and Getz, G.S., Steady state analysis of mitochondrial RNA after growth of yeast *Saccharomyces cerevisiae* under catabolite repression and derepression. *Journal of Biological Chemistry*, *1986*. ***261****(25): p. 11816-11822*.
111. Marykwas, D.L. and Fox, T.D., Control of the *Saccharomyces cerevisiae* regulatory gene PET494: transcriptional repression by glucose and translational induction by oxygen. *Molecular and Cellular Biology*, *1989*. ***9****(2): p. 484-491*.
112. Basu, U., Bostwick, A.M., Das, K., Dittenhafer-Reed, K.E., and Patel, S.S., Structure, mechanism, and regulation of mitochondrial DNA transcription initiation. *Journal of Biological Chemistry*, *2020*. ***295****(52): p. 18406-18425*.
113. Grivell, L.A., Nucleo-mitochondrial interactions in yeast mitochondrial biogenesis. *European Journal of Biochemistry*, *1989*. ***182****(3): p. 477-93*.
114. Turk, E.M., Das, V., Seibert, R.D., and Andrulis, E.D., The mitochondrial RNA landscape of *Saccharomyces cerevisiae*. *PLoS One*, *2013*. ***8****(10): p. e78105*.
115. Lambowitz, A.M. and Belfort, M., Introns as mobile genetic elements. *Annual review of biochemistry*, *1993*. ***62****: p. 587-622*.
116. Séraphin, B., Boulet, A., Simon, M., and Faye, G., Construction of a yeast strain devoid of mitochondrial introns and its use to screen nuclear genes involved in mitochondrial splicing. *Proceedings of the National Academy of Sciences*, *1987*. ***84****(19): p. 6810-6814*.

117. Wu, B., Buljic, A., and Hao, W., Extensive horizontal transfer and homologous recombination generate highly chimeric mitochondrial genomes in yeast. *Molecular Biology and Evolution*, *2015*. ***32****(10): p. 2559-70*.

118. Fonseca, P.L., De-Paula, R.B., Araújo, D.S., Tomé, L.M.R., Mendes-Pereira, T., Rodrigues, W.F.C., Del-Bem, L.-E., Aguiar, E.R., and Góes-Neto, A., Global characterization of fungal mitogenomes: new insights on genomic diversity and dynamism of coding genes and accessory elements. *Frontiers in Microbiology*, *2021: p. 3733*.

119. Lang, B.F., Laforest, M.-J., and Burger, G., Mitochondrial introns: a critical view. *Trends in Genetics*, *2007*. ***23****(3): p. 119-125*.

120. Rudan, M., Dib, P.B., Musa, M., Kanunnikau, M., Sobočanec, S., Rueda, D., Warnecke, T., and Kriško, A., Normal mitochondrial function in *Saccharomyces cerevisiae* has become dependent on inefficient splicing. *Elife*, *2018*. ***7****: p. e35330*.

121. Wang, G.-S. and Cooper, T.A., Splicing in disease: disruption of the splicing code and the decoding machinery. *Nature Reviews Genetics*, *2007*. ***8****(10): p. 749-761*.

122. Miyakawa, I., Sando, N., Kawano, S., Nakamura, S., and Kuroiwa, T., Isolation of morphologically intact mitochondrial nucleoids from the yeast, *Saccharomyces cerevisiae*. *Journal of Cell Science*, *1987*. ***88****(4): p. 431-439*.

123. Freel, K.C., Friedrich, A., and Schacherer, J., Mitochondrial genome evolution in yeasts: an all-encompassing view. *FEMS yeast research*, *2015*. ***15****(4)*.

124. Chen, X.J. and Clark-Walker, G.D., Unveiling the mystery of mitochondrial DNA replication in yeasts. *Mitochondrion*, *2018*. ***38****: p. 17-22*.

125. Williamson, D., The curious history of yeast mitochondrial DNA. *Nature Reviews Genetics*, *2002*. ***3****(6): p. 475*.

126. Formaggioni, A., Luchetti, A., and Plazzi, F., Mitochondrial Genomic Landscape: A Portrait of the Mitochondrial Genome 40 Years after the First Complete Sequence. *Life (Basel)*, *2021*. ***11****(7)*.

127. Yang, E.J., Pernice, W.M., and Pon, L.A., A role for cell polarity in lifespan and mitochondrial quality control in the budding yeast *Saccharomyces cerevisiae*. *iScience*, *2022*. ***25****(3): p. 103957*.

128. Berger, K.H. and Yaffe, M.P., Mitochondrial DNA inheritance in *Saccharomyces cerevisiae*. *Trends in Microbiology*, *2000*. ***8****(11): p. 508-513*.

129. Ling, F., Mikawa, T., and Shibata, T., Enlightenment of yeast mitochondrial homoplasmy: Diversified roles of gene conversion. *Genes*, *2011*. ***2****(1): p. 169-190*.

130. Sena, E.P., Revet, B., and Moustacchi, E., *In vivo* homologous recombination intermediates of yeast mitochondrial DNA analyzed by electron microscopy. *Molecular and General Genetics MGG*, *1986*. ***202****(3): p. 421-428*.

131. Stein, A., Kalifa, L., and Sia, E.A., Members of the RAD52 epistasis group contribute to mitochondrial homologous recombination and double-strand break repair in *Saccharomyces cerevisiae*. *PLoS genetics*, *2015*. ***11****(11): p. e1005664*.

132. Chen, X.J., Mechanism of homologous recombination and implications for aging-related deletions in mitochondrial DNA. *Microbiology and Molecular Biology Reviews*, *2013*. ***77****(3): p. 476-496*.

133. Valente, W.J., Ericson, N.G., Long, A.S., White, P.A., Marchetti, F., and Bielas, J.H., Mitochondrial DNA exhibits resistance to induced point and deletion mutations. *Nucleic acids research*, *2016*. ***44****(18): p. 8513-8524*.

134. King, M.P. and Attardi, G., Human cells lacking mtDNA: Repopulation with exogenous mitochondria by complementation. *Science*, *1989*. ***246****(4929): p. 500-503*.

135. Silva-Pinheiro, P. and Minczuk, M., The potential of mitochondrial genome engineering. *Nature Reviews Genetics*, *2022*. ***23****(4): p. 199-214*.

136. Pereira, C.V. and Moraes, C.T., Current strategies towards therapeutic manipulation of mtDNA heteroplasmy. *Front Biosci (Landmark Ed)*, *2017*. ***22****(6): p. 991-1010*.

137. Baile, M.G. and Claypool, S.M., The power of yeast to model diseases of the powerhouse of the cell. *Frontiers in bioscience (Landmark edition)*, *2013*. ***18****: p. 241*.

138. Pfeiffer, T. and Morley, A., An evolutionary perspective on the Crabtree effect. *Frontiers in Molecular Biosciences*, *2014*. ***1****: p. 17*.

139. Larosa, V. and Remacle, C., Transformation of the mitochondrial genome. *International Journal of Developmental Biology*, *2013*. ***57****(6-7-8): p. 659-665*.

140. Butow, R.A., Henke, R.M., Moran, J.V., Belcher, S.M., and Perlman, P.S., Transformation of *Saccharomyces cerevisiae* mitochondria using the biolistic gun, in Methods in enzymology. 1996, *Academic Press*. p. 265-278.

141. Suhm, T., Habernig, L., Rzepka, M., Kaimal, J.M., Andreasson, C., Buttner, S., and Ott, M., A novel system to monitor mitochondrial translation in yeast. *Microb Cell*, *2018*. ***5****(3): p. 158-164*.

142. Rzepka, M., Suhm, T., and Ott, M., Incorporation of reporter genes into mitochondrial DNA in budding yeast. *STAR protocols*, *2022*. ***3****(2): p. 101359*.

143. Mireau, H., Arnal, N., and Fox, T.D., Expression of Barstar as a selectable marker in yeast mitochondria. *Molecular Genetics and Genomics*, *2003*. ***270****(1): p. 1-8*.

144. Jauniaux, J.C., Urrestarazu, L.A., and Wiame, J.M., Arginine metabolism in *Saccharomyces cerevisiae*: subcellular localization of the enzymes. *Journal of bacteriology*, *1978*. ***133****(3): p. 1096-1107*.

145. Steele, D.F., Butler, C.A., and Fox, T.D., Expression of a recoded nuclear gene inserted into yeast mitochondrial DNA is limited by mRNA-specific translational activation. *Proceedings of the National Academy of Sciences*, *1996*. ***93****(11): p. 5253-5257*.

146. Mojica, F.J.M., Díez-Villaseñor, C., Soria, E., and Juez, G., Biological significance of a family of regularly spaced repeats in the genomes of Archaea, Bacteria and mitochondria. *Molecular microbiology*, *2000*. ***36****(1): p. 244-246*.

147. Barrangou, R., Fremaux, C., Deveau, H., Richards, M., Boyaval, P., Moineau, S., Romero, D.A., and Horvath, P., CRISPR provides acquired resistance against viruses in prokaryotes. *Science*, *2007*. ***315****(5819): p. 1709-1712*.

148. Jinek, M., Chylinski, K., Fonfara, I., Hauer, M., Doudna, J.A., and Charpentier, E., A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science*, *2012*. ***337****(6096): p. 816-821*.

149. Cong, L., Ran, F.A., Cox, D., Lin, S., Barretto, R., Habib, N., Hsu, P.D., Wu, X., Jiang, W., Marraffini, L.A., and Zhang, F., Multiplex genome engineering using CRISPR/Cas systems. *Science*, *2013*. ***339****(6121): p. 819-23*.

150. Adli, M., The CRISPR tool kit for genome editing and beyond. *Nature communications*, *2018*. ***9****(1): p. 1911*.

151. van der Oost, J. and Patinios, C., The genome editing revolution. *Trends in Biotechnology*, *2023*.

152. Krejci, L., Altmannova, V., Spirek, M., and Zhao, X., Homologous recombination and its regulation. *Nucleic acids research*, *2012*. ***40****(13): p. 5795-5818*.

153. Straub, S.P., Stiller, S.B., Wiedemann, N., and Pfanner, N., Dynamic organization of the mitochondrial protein import machinery. *Biological Chemistry*, *2016*. ***397****(11): p. 1097-1114*.

154. Wiedemann, N. and Pfanner, N., Mitochondrial machineries for protein import and assembly. *Annual review of biochemistry*, *2017*. ***86****(1): p. 685-714*.

155. Moulin, C., Caumont-Sarcos, A., and Ieva, R., Mitochondrial presequence import: Multiple regulatory knobs fine-tune mitochondrial biogenesis and homeostasis. *Biochimica et Biophysica Acta (BBA) - Molecular Cell Research*, *2019*. ***1866****(5): p. 930-944*.

156. Pfanner, N., Warscheid, B., and Wiedemann, N., Mitochondrial proteins: from biogenesis to functional networks. *Nature Reviews Molecular Cell Biology*, *2019*. ***20****(5): p. 267-284*.

157. Yang, X., Jiang, J., Li, Z., Liang, J., and Xiang, Y., Strategies for mitochondrial gene editing. *Computational and Structural Biotechnology Journal*, *2021*. ***19***: *p. 3319-3329*.

158. Tanaka, M., Borgeld, H.J., Zhang, J., Muramatsu, S., Gong, J.S., Yoneda, M., Maruyama, W., Naoi, M., Ibi, T., Sahashi, K.,... Yagi, K., Gene therapy for mitochondrial disease by delivering restriction endonuclease *Sma*I into mitochondria. *Journal of Biomedical Science*, *2002*. ***9***(6): *p. 534-541*.

159. Reddy, P., Ocampo, A., Suzuki, K., Luo, J., Bacman, Sandra R., Williams, Sion L., Sugawara, A., Okamura, D., Tsunekawa, Y., Wu, J.,... Izpisua Belmonte, Juan C., Selective elimination of mitochondrial mutations in the germline by genome editing. *Cell*, *2015*. ***161***(3): *p. 459-469*.

160. Bacman, S.R., Williams, S.L., Pinto, M., and Moraes, C.T., The use of mitochondria-targeted endonucleases to manipulate mtDNA, in Methods in enzymology, A.N. Murphy and D.C. Chan, Editors. 2014, *Academic Press*. p. 373-397.

161. Gammage, P.A., Rorbach, J., Vincent, A.I., Rebar, E.J., and Minczuk, M., Mitochondrially targeted ZFNs for selective degradation of pathogenic mitochondrial genomes bearing large-scale deletions or point mutations. *EMBO Molecular Medicine*, *2014*. ***6***(4): *p. 458-466*.

162. Minczuk, M., Papworth, M.A., Miller, J.C., Murphy, M.P., and Klug, A., Development of a single-chain, quasi-dimeric zinc-finger nuclease for the selective degradation of mutated human mitochondrial DNA. *Nucleic acids research*, *2008*. ***36***(12): *p. 3926-3938*.

163. Mok, B.Y., de Moraes, M.H., Zeng, J., Bosch, D.E., Kotrys, A.V., Raguram, A., Hsu, F., Radey, M.C., Peterson, S.B., Mootha, V.K., Mougous, J.D., and Liu, D.R., A bacterial cytidine deaminase toxin enables CRISPR-free mitochondrial base editing. *Nature*, *2020*. ***583***(7817): *p. 631-637*.

164. Cho, S.-I., Lee, S., Mok, Y.G., Lim, K., Lee, J., Lee, J.M., Chung, E., and Kim, J.-S., Targeted A-to-G base editing in human mitochondrial DNA with programmable deaminases. *Cell*, *2022*. ***185***(10): *p. 1764-1776. e12*.

165. Gammage, P.A., Moraes, C.T., and Minczuk, M., Mitochondrial genome engineering: the revolution may not be CRISPR-Ized. *Trends in Genetics*, *2018*. ***34***(2): *p. 101-110*.

166. Barrera-Paez, J.D. and Moraes, C.T., Mitochondrial genome engineering coming-of-age. *Trends in Genetics*, *2022*. ***38***(8): *p. 869-880*.

167. Weber-Lotfi, F., Ibrahim, N., Boesch, P., Cosset, A., Konstantinov, Y., Lightowlers, R.N., and Dietrich, A., Developing a genetic approach to investigate the mechanism of mitochondrial competence for DNA import. *Biochimica et Biophysica Acta (BBA)-Bioenergetics*, *2009*. ***1787***(5): *p. 320-327*.

168. Niazi, A.K., Mileshina, D., Cosset, A., Val, R., Weber-Lotfi, F., and Dietrich, A., Targeting nucleic acids into mitochondria: progress and prospects. *Mitochondrion*, *2013*. ***13***(5): *p. 548-558*.

169. Yamada, Y. and Harashima, H., Delivery of bioactive molecules to the mitochondrial genome using a membrane-fusing, liposome-based carrier, DF-MITO-Porter. *Biomaterials*, *2012*. ***33***(5): *p. 1589-1595*.

170. D'Souza, G.G., Rammohan, R., Cheng, S.-M., Torchilin, V.P., and Weissig, V., DQAsome-mediated delivery of plasmid DNA toward mitochondria in living cells. *Journal of controlled release*, *2003*. ***92***(1-2): *p. 189-197*.

171. Yu, H., Koilkonda, R.D., Chou, T.-H., Porciatti, V., Ozdemir, S.S., Chiodo, V., Boye, S.L., Boye, S.E., Hauswirth, W.W., and Lewin, A.S., Gene delivery to mitochondria by targeting modified adenoassociated virus suppresses Leber's hereditary optic neuropathy in a mouse model. *Proceedings of the National Academy of Sciences*, *2012*. ***109***(20): *p. E1238-E1247*.

172. Lee, M., Choi, J.S., and Ko, K.S., Mitochondria targeting delivery of nucleic acids. *Expert Opinion on Drug Delivery*, *2008*. ***5***(8): *p. 879-887*.

173. Weissig, V. and Torchilin, V.P., Cationic bolasomes with delocalized charge centers as mitochondria-specific DNA delivery systems. *Advanced Drug Delivery Reviews*, *2001*. ***49***(1): *p. 127-149*.

174. Yuan, P., Mao, X., Wu, X., Liew, S.S., Li, L., and Yao, S.Q., Mitochondria-targeting, intracellular delivery of native proteins using biodegradable silica nanoparticles. *Angewandte Chemie International Edition*, *2019*. ***58***(23): *p. 7657-7661*.

175. Rinehart, J., Krett, B., Rubio, M.A.T., Alfonzo, J.D., Söll, D.J.G., and development, *Saccharomyces cerevisiae* imports the cytosolic pathway for Gln-tRNA synthesis into the mitochondrion. *2005*. ***19***(5): *p. 583-592*.

176. Tarassov, I.A. and Entelis, N.S., Mitochondrially-imported cytoplasmic tRNA Lys (CUU) of *Saccharomyces cerevisiae*: *in vivo* and *in vitro* targetting systems. *Nucleic acids research*, *1992*. ***20***(6): *p. 1277-1281*.

177. Li, K., Smagula, C., Parsons, W., Richardson, J., Gonzalez, M., Hagler, H., and Williams, R., Subcellular partitioning of MRP RNA assessed by ultrastructural and biochemical analysis. *Journal of Cell Biology*, *1994*. ***124***(6): *p. 871-882*.

178. Doersen, C.J., Guerrier-Takada, C., Altman, S., and Attardi, G., Characterization of an RNase P activity from HeLa cell mitochondria. Comparison with the cytosol RNase P activity. *Journal of Biological Chemistry*, *1985*. ***260***(10): *p. 5942-9*.

179. Magalhães, P.J., Andreu, A.L., and Schon, E.A., Evidence for the presence of 5S rRNA in mammalian mitochondria. *Molecular Biology of the Cell*, *1998*. ***9***(9): *p. 2375-82*.

180. Rubio, M.A., Rinehart, J.J., Krett, B., Duvezin-Caubet, S., Reichert, A.S., Söll, D., and Alfonzo, J.D., Mammalian mitochondria have the innate ability to import tRNAs by a mechanism distinct from protein import. *Proceedings of the National Academy of Sciences*, *2008*. ***105***(27): *p. 9186-91*.

181. Schneider, A., Mitochondrial tRNA import and its consequences for mitochondrial translation. *Annual review of biochemistry*, *2011*. ***80***: *p. 1033-1053*.

182. Tan, T.H.P., Pach, R., Crausaz, A., Ivens, A., and Schneider, A., tRNAs in *Trypanosoma brucei*: Genomic organization, expression, and mitochondrial import. *Molecular and Cellular Biology*, *2002*. ***22***(11): *p. 3707-3717*.

183. Entelis, N., Brandina, I., Kamenski, P., Krasheninnikov, I.A., Martin, R.P., and Tarassov, I., A glycolytic enzyme, enolase, is recruited as a cofactor of tRNA targeting toward mitochondria in *Saccharomyces cerevisiae*. *Genes & development*, *2006*. ***20***(12): *p. 1609-1620*.

184. Tarassov, I., Entelis, N., and Martin, R., Mitochondrial import of a cytoplasmic lysine-tRNA in yeast is mediated by cooperation of cytoplasmic and mitochondrial lysyl-tRNA synthetases. *The EMBO journal*, *1995*. ***14***(14): *p. 3461-3471*.

185. Tarassov, I., Entelis, N., and Martin, R.P., An intact protein translocating machinery is required for mitochondrial import of a yeast cytoplasmic tRNA. *J Mol Biol*, *1995*. ***245***(4): *p. 315-23*.

186. Kamenski, P., Smirnova, E., Kolesnikova, O., Krasheninnikov, I.A., Martin, R.P., Entelis, N., and Tarassov, I.J.M., tRNA mitochondrial import in yeast: Mapping of the import determinants in the carrier protein, the precursor of mitochondrial lysyl-tRNA synthetase. *Mitochondrion*, *2010*. ***10***(3): *p. 284-293*.

187. Wang, G., Chen, H.-W., Oktay, Y., Zhang, J., Allen, E.L., Smith, G.M., Fan, K.C., Hong, J.S., French, S.W., and McCaffery, J.M., PNPase regulates RNA import into mitochondria. *Cell*, *2010*. ***142***(3): *p. 456-467*.

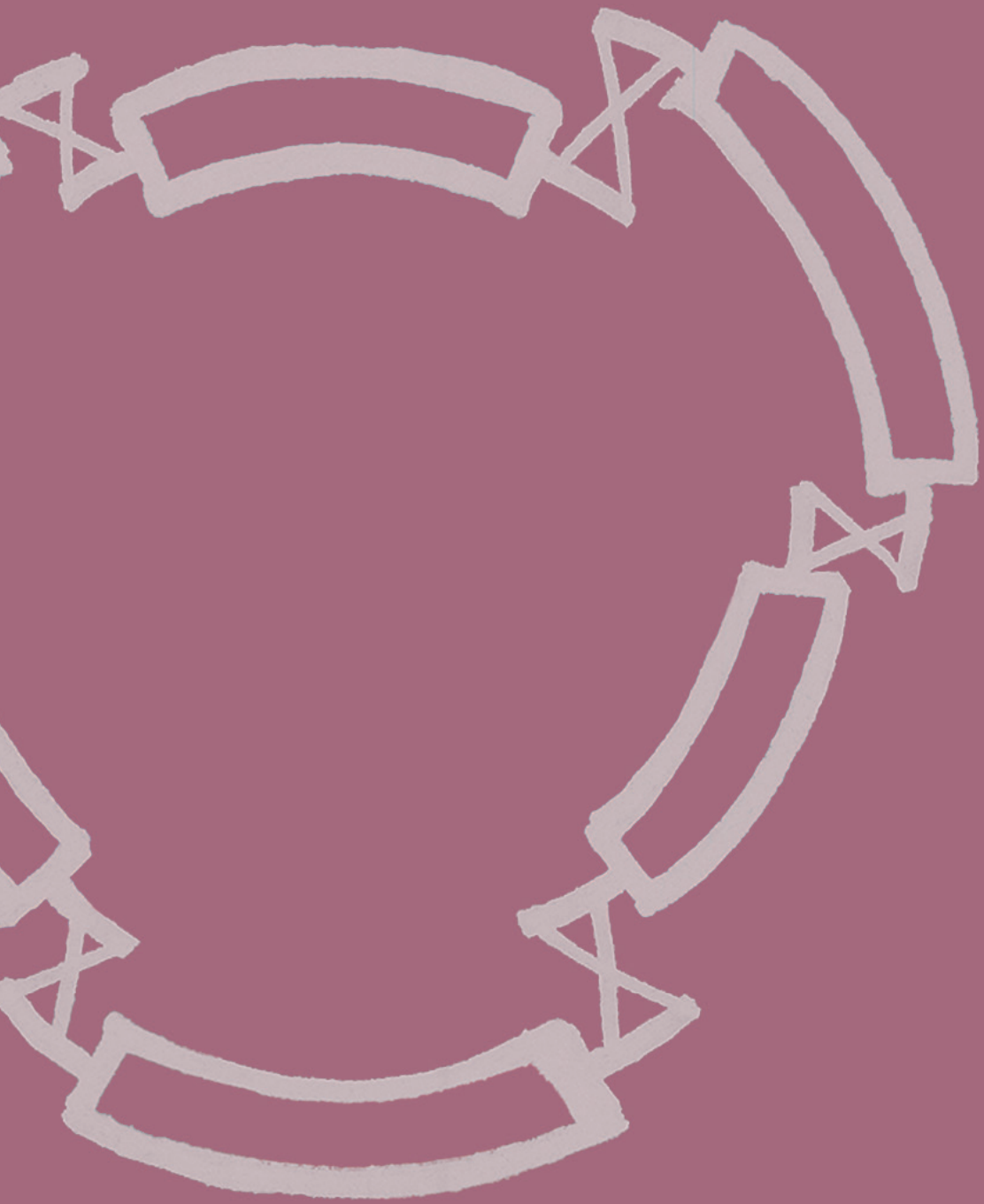
188. Wang, G., Shimada, E., Zhang, J., Hong, J.S., Smith, G.M., Teitell, M.A., and Koehler, C.M., Correcting human mitochondrial mutations with targeted RNA import. *Proceedings of the National Academy of Sciences*, *2012*. ***109***(13): *p. 4840-4845*.

189. Smirnov, A., Comte, C., Mager-Heckel, A.M., Addis, V., Krasheninnikov, I.A., Martin, R.P., Entelis, N., and Tarassov, I., Mitochondrial enzyme rhodanese is essential for 5 S ribosomal RNA import into human mitochondria. *Journal of Biological Chemistry*, *2010*. ***285***(40): *p. 30792-803*.

190. Sharon, E., Chen, S.-A.A., Khosla, N.M., Smith, J.D., Pritchard, J.K., and Fraser, H.B., Functional genetic variants revealed by

massively parallel precise genome editing. *Cell*, *2018*. ***175****(2): p. 544-557. e16*.

191. Chandramouly, G., Zhao, J., McDevitt, S., Rusanov, T., Hoang, T., Borisonnik, N., Treddinick, T., Lopezcolorado, F.W., Kent, T., Siddique, L.A.,... Pomerantz, R.T., Polθ; reverse transcribes RNA and promotes RNA-templated DNA repair. *Science Advances*, *2021*. ***7****(24): p. eabf1771*.
192. Storici, F., Bebenek, K., Kunkel, T.A., Gordenin, D.A., and Resnick, M.A., RNA-templated DNA repair. *Nature*, *2007*. ***447****(7142): p. 338-341*.
193. Kolesnikova, O., Entelis, N., Kazakova, H., Brandina, I., Martin, R.P., and Tarassov, I., Targeting of tRNA into yeast and human mitochondria: the role of anticodon nucleotides. *Mitochondrion*, *2002*. ***2****(1-2): p. 95-107*.
194. Kolesnikova, O., Kazakova, H., Comte, C., Steinberg, S., Kamenski, P., Martin, R.P., Tarassov, I., and Entelis, N., Selection of RNA aptamers imported into yeast and human mitochondria. *RNA*, *2010*. ***16****(5): p. 926-41*.
195. Loutre, R., Heckel, A.M., Smirnova, A., Entelis, N., and Tarassov, I., Can mitochondrial DNA be CRISPRized: Pro and contra. *IUBMB life*, *2018*. ***70****(12): p. 1233-1239*.
196. Zelenka, J., Alán, L., Jabůrek, M., and Ježek, P., Import of desired nucleic acid sequences using addressing motif of mitochondrial ribosomal 5S-rRNA for fluorescent *in vivo* hybridization of mitochondrial DNA and RNA. *Journal of bioenergetics and biomembranes*, *2014*. ***46****(2): p. 147-156*.
197. Weber-Lotfi, F. and Dietrich, A., Targeting therapeutic nucleic acids into mitochondria: A long challenge, in Mitochondrial biology and experimental therapeutics. 2018, *Springer*. p. 565-592.
198. Jo, A., Ham, S., Lee, G.H., Lee, Y.I., Kim, S., Lee, Y.S., Shin, J.H., and Lee, Y., Efficient mitochondrial genome editing by CRISPR/Cas9. *BioMed Research International*, *2015*. ***2015****: p. 305716*.
199. Hussain, S.-R.A., Yalvac, M.E., Khoo, B., Eckardt, S., and McLaughlin, K.J., Adapting CRISPR/Cas9 system for targeting mitochondrial genome. 2020, Cold Spring Harbor Laboratory.
200. Wang, B., Lv, X., Wang, Y., Wang, Z., Liu, Q., Lu, B., Liu, Y., and Gu, F., CRISPR/Cas9-mediated mutagenesis at microhomologous regions of human mitochondrial genome. *Science China Life Sciences*, *2021*. ***64****(9): p. 1463-1472*.
201. Amai, T., Tsuji, T., Ueda, M., and Kuroda, K., Development of a mito-CRISPR system for generating mitochondrial DNA-deleted strain in *Saccharomyces cerevisiae*. *Bioscience*, *Biotechnology*, *and Biochemistry*, *2020*. ***85****(4): p. 895-901*.
202. Bi, R., Li, Y., Xu, M., Zheng, Q., Zhang, D.-F., Li, X., Ma, G., Xiang, B., Zhu, X., Zhao, H., Huang, X., Zheng, P., and Yao, Y.-G., Direct evidence of CRISPR-Cas9-mediated mitochondrial genome editing. *The Innovation*, *2022*. ***3****(6): p. 100329*.
203. Bian, W.-P., Chen, Y.-L., Luo, J.-J., Wang, C., Xie, S.-L., and Pei, D.-S., Knock-in strategy for editing human and zebrafish mitochondrial DNA using Mito-CRISPR/Cas9 system. *ACS synthetic biology*, *2019*. ***8****(4): p. 621-632*.
204. Schmiderer, L., Yudovich, D., Oburoglu, L., Hjort, M., and Larsson, J., Site-specific CRISPR-based mitochondrial DNA manipulation is limited by gRNA import. *Scientific reports*, *2022*. ***12****(1): p. 18687*.
205. Tang, J.-X., Pyle, A., Taylor, R.W., and Oláhová, M., Interrogating mitochondrial biology and disease using CRISPR/Cas9 gene editing. *Genes*, *2021*. ***12****(10): p. 1604*.
206. Antón, Z., Mullally, G., Ford, H.C., Van Der Kamp, M.W., Szczelkun, M.D., and Lane, J.D., Mitochondrial import, health and mtDNA copy number variability seen when using type II and type V CRISPR effectors. *Journal of Cell Science*, *2020*. ***133****(18): p. jcs248468*.



CHAPTER 2
SYNTHETIC GENOMICS FROM A YEAST PERSPECTIVE

Charlotte C. Koster*, Eline D. Postma*, Ewout Knibbe*, Céline Cleij*, Pascale Daran-Lapujade
*These authors have contributed equally to this work and share first authorship

This chapter was published in Frontiers in Biotechnology and Bioengineering
doi: 10.3389/fbioe.2022.869486

Abstract

Synthetic Genomics focuses on the construction of rationally designed chromosomes and genomes and offers novel approaches to study biology and to construct synthetic cell factories. Currently, progress in Synthetic Genomics is hindered by the inability to synthesize DNA molecules longer than a few hundred base pairs, while the size of the smallest genome of a self-replicating cell is several hundred thousand base pairs. Methods to assemble small fragments of DNA into large molecules are therefore required. Remarkably powerful at assembling DNA molecules, the unicellular eukaryote *Saccharomyces cerevisiae* has been pivotal in the establishment of Synthetic Genomics. Instrumental in the assembly of entire genomes of various organisms in the past decade, the *S. cerevisiae* genome foundry has a key role to play in future Synthetic Genomics developments.

2.1 Introduction

Synthetic Genomics (SG) is a recent synthetic biology discipline that focuses on the construction of rationally designed chromosomes and genomes. SG offers a novel approach to address fundamental biological questions by restructuring, recoding, and minimizing (parts of) genomes (as recently reviewed by [1]). SG is now spurring technological developments in academia and has a strong future potential in industry [2, 3]). Humankind's best microbial friend, the baker's yeast *Saccharomyces cerevisiae*, has played, and continues to play a key role in SG advances, both by enabling the construction of chromosomes for other hosts, and in the refactoring of its own genome. This mini review explores the reasons for this strategic positioning of *S. cerevisiae* in SG, surveys the main achievements enabled by this yeast and reflects on future developments.

2.2 Synthetic genomics from a yeast perspective

Current limitations of genome assembly

While small-sized viral chromosomes were the first to be chemically synthetized, the breakthrough in the field of SG came with the synthesis and assembly of the 592 kilobase (kb) chromosome of *Mycoplasma genitalium* [4, 5]. The unicellular eukaryote *Saccharomyces cerevisiae* has made a key contribution to this famous milestone. To understand how this microbe, commonly used in food and beverages, contributes to the assembly of synthetic genomes, let us recapitulate how synthetic chromosomes can be constructed (Figure 2.1).

It starts with the customized synthesis of short DNA molecules called oligonucleotides. Oligonucleotides are mostly synthetized via phosphoramidite chemistry [6], a 40-year-old method that, despite decades of technological developments, struggles to deliver error-free oligonucleotides longer than 200 base pairs (bp). While the implementation of microarrays has substantially decreased the synthesis cost, it has not increased oligo length, an achievement that requires new synthesis methods [7]. Enzymatic alternatives for DNA synthesis are under development [8, 9], but still have considerable shortcomings regarding automation and scalability that must be overcome before commercial scale can be considered (reviewed in [10-13]). Considering that a theoretical minimal genome would be around 113 kb long [14] and that the first fully synthesized genome of *M. genitalium* contains 583 kb [4], thousands of oligos must be stitched together to construct a complete synthetic genome. These DNA oligos can be assembled into longer DNA fragments owing to a plethora of *in vitro* methods (reviewed in [11, 15, 16]). A method that has gained tremendous popularity since its development is the homology-based Gibson isothermal assembly [17], devised to assemble the *M. genitalium* genome. As all *in vitro* methods, Gibson assembly is limited by the number of fragments that can reliably be stitched together in one reaction, usually around a dozen, requiring a stepwise assembly procedure of increasingly large genomic DNA constructs [18]. DNA must be recovered from the reaction, amplified and verified in each round, to allow further processing. Selection and amplification of

correctly cloned DNA is routinely performed in *Escherichia coli*, however, maintenance of large constructs of exogenous DNA, especially from prokaryotic origins, in this bacterium is often limited by expression and toxicity of gene products [19]. *In vitro* alternatives for efficient and faithful selection and amplification of correctly assembled DNA are under development, but these are currently limited in length of amplified DNA and scalability [20-23]. While in principle stepwise *in vitro* assembly can lead to a DNA molecule of any size, and selection and amplification in *E. coli* worked well for DNA constructs up to 72 kb, *E. coli* had great difficulties maintaining quarter *M. genitalium* genomes, causing Gibson and colleagues to turn to baker's yeast [4, 5].

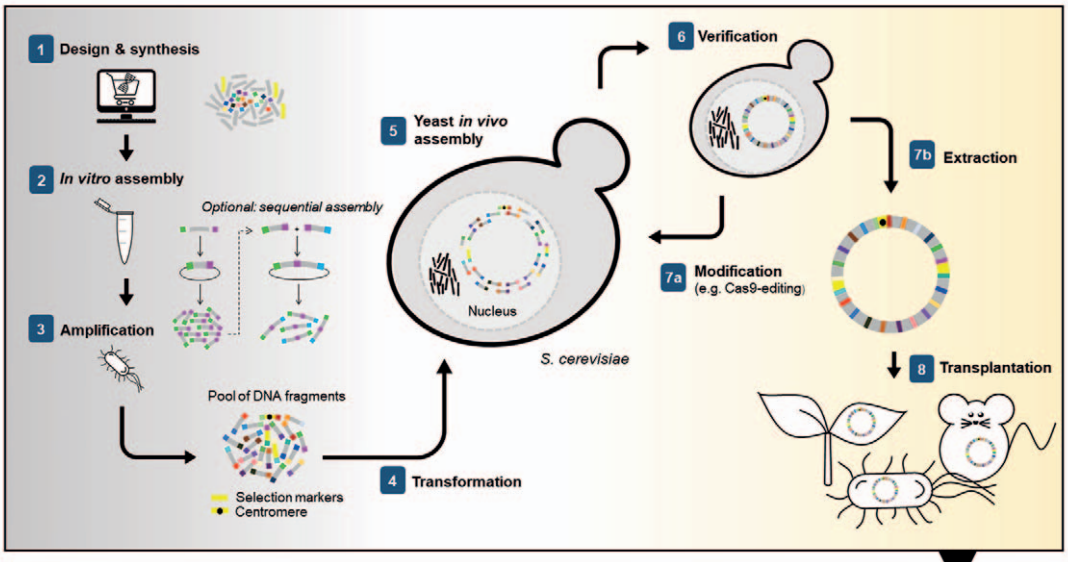
Saccharomyces cerevisiae as a genome foundry

S. cerevisiae seems a logical host for SG as it naturally maintains a 12 Mb genome consisting of 16 chromosomes ranging from 230 to 1500 kb in its haploid version, lives as polyploid in natural environments, and is extremely robust to changes in genome content and architecture [24]. The extreme robustness of *S. cerevisiae* to supernumerary, chimeric chromosomes, a key feature for SG, was already demonstrated in the late '80s [25, 26]. A second key feature of *S. cerevisiae* is its preference for homologous recombination (HR) to repair double-strand DNA breaks [27], a rare trait among eukaryotes. *S. cerevisiae* ability to efficiently and with high fidelity stitch together linear DNA molecules that present homologous regions as short as 40 bp [28] at their ends, was rapidly valorized for genetic manipulations and assembly of heterologous DNA. Recently renamed *in vivo* assembly, this cloning technique (Figure 2.1) contributes to the remarkable genetic tractability and popularity of *S. cerevisiae* as model and industrial microbe [17, 29]. The combination of *S. cerevisiae*'s HR efficiency and fidelity, chromosome maintenance and propagation enabled the construction of the full *Mycoplasma* genome. Reflecting that "*in the future, it may be advantageous to make greater use of yeast recombination to assemble chromosomes*", this study propelled *S. cerevisiae* as powerful 'genome foundry' [4]. In the challenge to synthesize genomes, Ostrov and colleagues rightfully identified assembly of these long DNA constructs as '*the most critical hurdle*' [10]. To date, *S. cerevisiae* has been key to assembling entire or partial genomes in most synthetic genome projects (Table 2.1). For instance, the entire 785 kb refactored *Caulobacter crescentus* (renamed *C. ethensis*) genome was assembled *in vivo* from 16 fragments [30], while the recoded *E. coli* genome was split over 10 fragments of 91 to 136 kb individually assembled in yeast, and then sequentially integrated in the *E. coli* chromosome to replace native segments [31] (Table 2.1). *In vivo* assembly also proved to be powerful in assembling and modifying genomes of organisms that are poorly amenable to genome editing; the rapid and faithful HR-based assembly of *S. cerevisiae* recently enabled the reconstruction of a synthetic SARS-CoV-2 genome in a single week [32], and has been shown to be a promising host for *in vivo* assembly and modification of other viral genomes [33] as well as the genomes of various pathogens [34] and even a 101 kb human gene, which was transplanted into mouse embryonic cells [35] (Table 2.1). Moreover, *S. cerevisiae* was selected for the construction of the first synthetic eukaryotic genome. The international Sc2.0 consortium, spearheaded by Jef Boeke, undertook less than ten years ago the daunting task of synthesizing recoded versions of the 16 yeast chromosomes. Via stepwise, systematic replacement of 30 to 40 kb (using *ca.* 12 DNA fragments of 2 to 4 kb) of the native yeast sequence, the consortium is close to the completion of the largest synthetic genome to date [36, 37], with the ambition to reshape and minimize the *S. cerevisiae* genome [38].

While *S. cerevisiae* is not the only microbial host available for the construction of (neo)chromosomes (Figure 2.1), several key features make it superior to its bacterial alternatives *Bacillus subtilis* and *E. coli* as genome foundry: i) *S. cerevisiae* has the natural ability to carry large amounts of DNA and therefore to host multiple exogenous bacterial genomes [34]; ii) *E. coli* frequently struggles with toxicity caused by the expression of exogenous bacterial sequences [5, 19, 39], while *S. cerevisiae* is very robust to the presence of heterologous DNA from prokaryotic or eukaryotic origin [40]; iii) *S. cerevisiae* can, in a single transformation, assemble many DNA oligonucleotides into (partial) genomes. *B. subtilis* can also maintain large exogenous DNA constructs, but requires a stepwise method for DNA assembly, in which each DNA part is integrated sequentially into *B. subtilis* genome [41]. This approach is intrinsically more labor-intensive and time-consuming than *S. cerevisiae* single transformation assembly.
Surprised by *S. cerevisiae* genetic tractability, Gibson and colleagues wondered "*how many pieces can be assembled in yeast in a single step?*" [4]. Pioneering a SG approach for metabolic engineering based on

modular, specialized synthetic chromosomes, Postma *et al.* probed this limit recently in our lab by constructing 100 kb artificial linear and circular neochromosomes from 44 DNA parts in a single transformation [42, 43]. The remarkable efficiency of *in vivo* assembly (36 % of assemblies faithful to design) revealed that its limit has clearly not been reached yet, and that future systematic studies are required to evaluate the true potential of *S. cerevisiae* as a genome foundry. The supernumerary chromosomes were shown to stably maintain complete heterologous pathways as well as the yeast's central carbon metabolism, underlining the potential of yeast synthetic genomics in the development of optimized cell-factories. Once assembled, synthetic chromosomes could be easily edited in *S. cerevisiae* thanks to its efficient HR and rich molecular toolbox.

A



B

	In vitro	*E. coli*	*B. subtilis*[1]	*S. cerevisiae*
One-pot assembly				
Number of fragments	Up to 20 (Chandran, S., 2017)	Up to 6 (Kostylev, M. *et al.*, 2015)	2 (Itaya, M. *et al.*, 2008)	Up to 44 (Postma, E.D. *et al.*, 2021)
Construct size	Up to 144 kb (Gibson, D.G. *et al.*, 2008)	Up to 180 kb (Kostylev, M. *et al.*, 2015)	Up to 200 kb (Itaya, M. *et al.*, 2018)	Up to 1.66 Mb (Tagwerker, C. *et al.*, 2012)
Sequential assembly				
Number of assembled fragments	600 (Gibson, D.G. *et al.*, 2010b)[2]	-	31 (Itaya, M. *et al.*, 2008)	1078 (Gibson, D.G. *et al.*, 2010a)
Number of sequential HR events	4	-	18	3
Final construct size	16.3 kb	-	134.5 kb	1.08 Mb
Other factors				
Compatibility with heterologous genes	***	*	***	***
DNA yield	*	***	Unknown	**
Overall proficiency to assemble synthetic genomes	-[3]	*	**	***

Figure 2.1. *In vivo* and *in vitro* approaches for DNA assembly in synthetic genomics. (A) Simplified overview of chromosome construction using *Saccharomyces cerevisiae* for genome assembly and production. (B) Strengths and weaknesses of *in vitro* and *in vivo* assembly methods. [(1)] Assembly of fragments in *B. subtilis* is performed by integration into the host genome. [(2)] Between rounds of sequential assembly, transformation into *E. coli* is conventional for selection and amplification of constructs. [(3)] Requires *in vivo* amplification and selection in a microbial host.

Challenges in genome assembly using yeast

While *S. cerevisiae* is natively proficient for SG, several aspects of *in vivo* assembly in yeast are still far from optimal. Firstly, compared to bacterial alternatives, *S. cerevisiae* cells grow slowly with a maximum specific growth rate around 0.4 - 0.5 h^{-1} and are hard to disrupt due to their sturdy cell wall. Considering that large DNA constructs above a few hundred kilobases are sensitive to shear stress, chromosome extraction and purification from *S. cerevisiae* is possible, but remains tenuous and inefficient, leading to low DNA yields and potentially damaged chromosomes [44]. Secondly, the strength of *S. cerevisiae* can become its weakness, as the HR machinery can be overzealous and recombine any (short) DNA sequence with homology within or between the (neo)chromosomes, which may lead to misassemblies. Lastly, non-homologous end joining and microhomology-mediated end joining, DNA repair mechanisms that assemble pieces of DNA with no or minimal homology, are present in *S. cerevisiae* with low activity [45, 46], and can also cause misassemblies. Similar to how *E. coli* was engineered to become a lab tool for DNA amplification, these shortcomings could be alleviated by engineering *S. cerevisiae* into a more powerful genome foundry.

Are there future alternatives to *S. cerevisiae*? Naturally, *B. subtilis* and *E. coli* could also be engineered. However, considering the minute fraction of the vast microbial biodiversity that has been tested for genetic accessibility and DNA assembly, it is likely that microbes yet to be discovered are even better genome foundries. Environments causing extreme DNA damage (high radiation, toxic chemicals, etc.) might be a source of HR-proficient organisms (e.g., [47, 48]) better suited for SG. In a more distant future, *in vitro* alternatives might replace the need for live DNA foundries altogether, thereby accelerating and simplifying genome construction. However, this will require major technological advances in *in vitro* DNA assembly and amplification. Already substantial efforts have led to the development of methods for DNA amplification, such as rolling circle amplification by the phage ϕ29 DNA polymerase [49, 50], recently implemented for the amplification of a 116 kb multipartite genome [20] and the *in vitro* amplification of synthetic genomes using the *E. coli* replisome, which already demonstrated to be capable of amplification of 1 Mb synthetic genomes [22]. Targets for improvement of these methods are the maximal length of amplified DNA fragments, the yield of amplification, the need for restriction of the amplified, concatenated molecules or the formation of non-specifically amplified products. The development of an *in vitro* approach that can parallel *S. cerevisiae in vivo* assembly capability seems even more challenging. While an interesting avenue might be to transplant *S. cerevisiae* HR DNA repair *in vitro*, it presents a daunting task considering that all players and their respective role have not been fully elucidated yet [46, 51]. Still, considering that highly complex systems such as the transcription and translation machineries have been successfully implemented *in vitro* and are commercially available [52], cell-free *S. cerevisiae* HR might become a reality in the coming years.

2.3 Outlook

Since the first genome synthesis in 2008, relatively few genomes have been synthetized. Low-cost, customizable construction of designer genomes, currently accessible for small viral, organellar or bacterial constructs, is still out of reach for large (eukaryotic) genomes. There are still numerous technical, financial, and computational hurdles that must be overcome on the road to microbial designer genomes, tailored to applications in bio-based industry. Here we reviewed why the yeast *S. cerevisiae i*s a key organism in the field of SG, however, the spectrum of available hosts is expected to increase as research in SG advances. For example, a recent study shows improving the HR capacity of the industrially relevant yeast *Yarrowia lipolytica* could greatly expand the potential applications of SG in bio-based processes [53].

In the near future, SG is anticipated to contribute to various fields, such as a platform technology for industrial biotechnological processes [3, 43], as a new means for data storage [54] and for the development of new cell therapies and other medical applications, which is the ambition of the Genome Project-Write [55]. In parallel, worldwide bottom-up approaches endeavor to construct synthetic cells from scratch, such as the European consortia BaSyC (http://www.basyc.nl), MaxSynBio (https://www.maxsynbio.mpg.de) and the Synthetic cell initiative (http://www.syntheticcell.eu) and the US-based Build-a-cell initiative (http://buildacell.io) (reviewed in [56]). Looking further ahead, SG may even assist in understanding and engineering entire ecosystems by assembly of a metagenomes in a single cell [57]. SG, albeit still in its infancy and mostly limited to academic research, has bright days ahead, and *S. cerevisiae* is foreseen to remain a valuable, if not indispensable, SG tool for the coming decade.

Table 2.1. Overview of the contribution of *S. cerevisiae* in synthetic genomics by the assembly of large (> 100 kb) DNA constructs.

	Donor DNA	Number of fragments[1]	Approximate size of fragments[1,2]	Approximate size of final construct	Aim of yeast assembly	Source
Viruses	Herpes simplex type 1	11	14 kb	152 kb	Assembly and modification of viral genome, transfection and reconstitution in mammalian cells.	[58]
	Autographa californica nucleopolyhedrovirus	4	45 kb	145 kb	Assembly and modification of viral genome, transfection and reconstitution in insect cells.	[59]
	Cytomegalovirus isolate Toledo	3	116 kb	230 kb	Assembly and modification of viral genome, transfection and reconstitution in mammalian cells.	[60]
Prokaryotes	*Mycoplasma genitalium*	6	Up to 144 kb	592 kb	Assembly of synthetic *M. genitalium* genome which could not be stably maintained in *E. coli.*	[4]
	Mycoplasma genitalium	25	17-35 kb	592 kb	Assembly of synthetic *M. genitalium* genome from short fragments, exploring assembly capacity in yeast.	[5]
	Mycoplasma mycoides	11	100 kb	1 Mb	Assembly of synthetic *M. mycoides* genome, transplantation to a recipient cell to create the first bacterial cell controlled by a synthesized genome.	[61]
	Mycoplasma pneumonia	2	10 kb - 816 kb	826 kb	Insertion of yeast regulatory elements in the full *M. pneumonia* genome to allow for cloning and engineering of the genome.	[34, 62]
	Mycoplasma hominis	2	5 kb – 665 kb	670 kb	Insertion of yeast regulatory elements in the full *M. hominis* genome to allow for cloning and engineering of the genome.	[63]
	Acholeplasma laidlawii	3[3]	121-897 kb	1.38 Mb	Exploring potential toxicity when assembling bacterial genomes in yeast.	[64]
	Escherichia coli	3	185 – 660 kb	1.03 Mb	Assembly of a minimal *E. coli* genome by Cas9-induced recombination of partial genomes.	[65]
	Escherichia coli	7 - 14	6-13 kb	100 kb	Assembly of recoded *E. coli* partial genomes, used to replace the *E. coli* genome by a recoded synthetic genome.	[31]
	Caulobacter crescentus	16	38-65 kb	785 kb	Assembly of a minimized and synthetic *C. crescentus* genome, recoded to be compatible with chemical DNA synthesis and transplanted in a recipient cell.	[30]
	Prechlorococcus marinus	2	580-675 kb	1.66 Mb	Exploring assembly capacity and DNA stability of exogenous genomes in yeast.	[40]
	Synechococcus elongatus	4	100 - 200 kb	454 kb	Exploring the ability to clone genomes with high G/C-content in yeast.	[66]
Algae	*Phaeodactylum tricornutum*	5	106-128 kb	497 kb	Assembly of DNA with a moderate G + C content as a case study for assembly and modification of eukaryotic chromosomes in yeast.	[67]
	Chlamydomonas reinhardtii chloroplast genome	6	34-129 kB	230 kb	Assembly of a partial *C. reinhardtii* chloroplast genome to create genetic diversity at multiple loci at once.	[68]

Yeasts	Yeast chromosome XII	33[4]	26-39 kb	976 kb	Assembly of a megabase synthetic yeast chromosome harboring the highly repetitive ribosomal DNA locus.	[69]
	Single-chromosome yeast	154	230-1500 kb	11 Mb	Assembly of all sixteen *S. cerevisiae* chromosomes into a single chromosome.	[24]
	Yeast neochromosome	44	2.5 kb	100 kb	Assembly of a circular supernumerary *S. cerevisiae* neochromosome that can act as a platform for modular genome engineering.	[42]
	Yeast neochromosome for pathway engineering	43	2.5 – 5 kb	100 kb	Assembly of a circular and linear supernumerary *S. cerevisiae* neochromosomes for expression of heterologous and essential metabolic pathways.	[43]
Other	Human *HPRT1* gene	13	3-83 kb	125 kb	Assembly of a synthetic human *HRPT1* gene and transplantation and expression in mammalian cells.	[35]
	Artificial data storage chromosome	5	40 kb	254 kb	Assembly of a *S. cerevisiae* artificial chromosome containing data-encoded DNA for digital data storage.	[54]

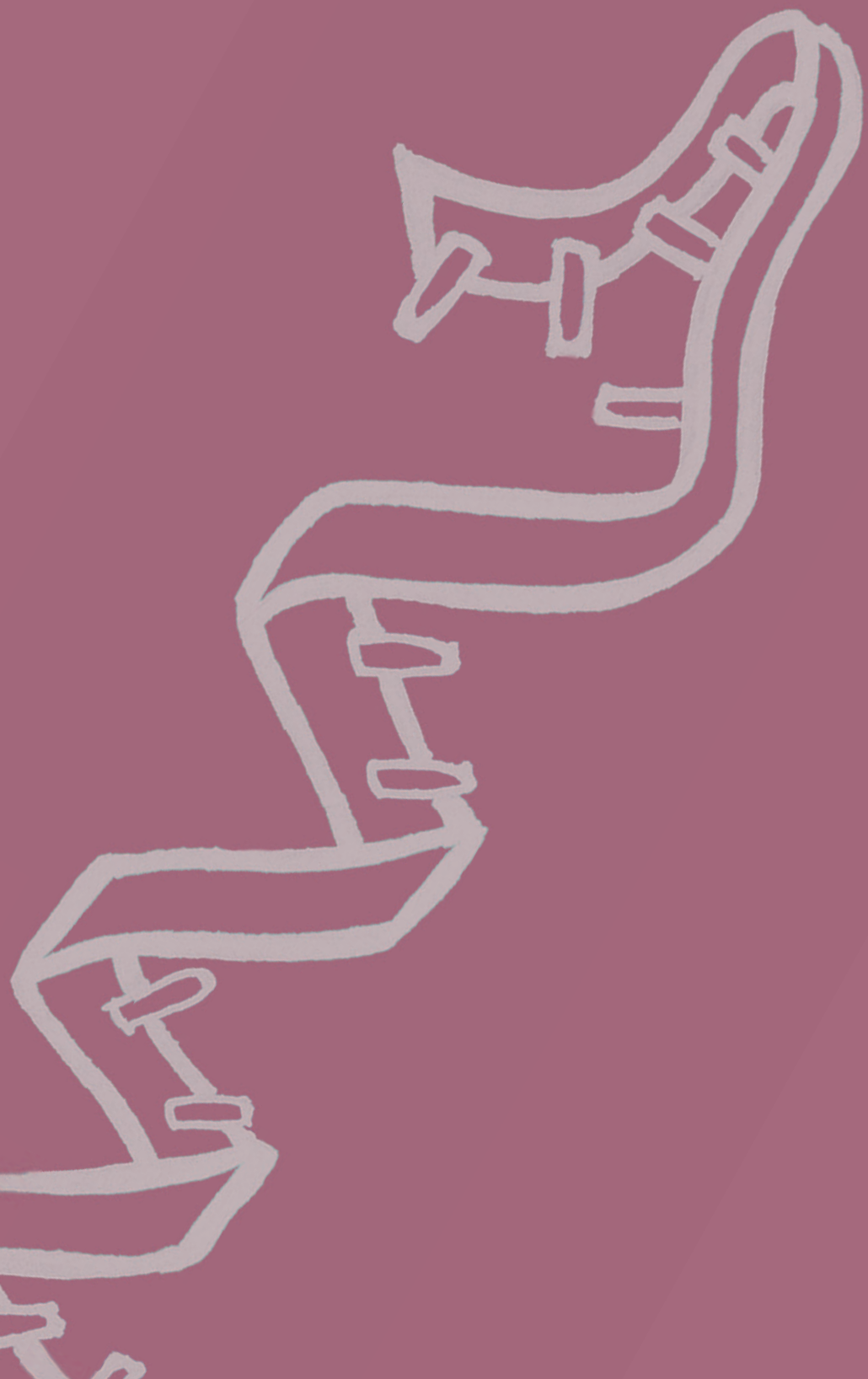
[1] In case of a sequential assembly, the fragment number and size of the last assembly is used
[2] Short backbones containing regulatory elements such as CEN/ARS and markers not included
[3] Initial assembly of the entire genome failed due to gene toxicity
[4] Assembly was performed by stepwise integration in multiple rounds

References

1. Coradini, A.L.V., Hull, C.B., and Ehrenreich, I.M., Building genomes to understand biology. *Nature Communications*, *2020*. ***11****(1)*.
2. Zhang, W., Mitchell, L.A., Bader, J.S., and Boeke, J.D., Synthetic genomes. *Annual Review of Biochemistry*, *2020*. ***89****: p. 77-101*.
3. Schindler, D., Genetic engineering and synthetic genomics in yeast to understand life and boost biotechnology. *Bioengineering*, *2020*. ***7****(4): p. 137*.
4. Gibson, D.G., Benders, G.A., Andrews-Pfannkoch, C., Denisova, E.A., Baden-Tillson, H., Zaveri, J., Stockwell, T.B., Brownley, A., Thomas, D.W., and Algire, M.A., Complete chemical synthesis, assembly, and cloning of a *Mycoplasma genitalium* genome. *Science*, *2008*. ***319****(5867): p. 1215-1220*.
5. Gibson, D.G., Benders, G.A., Axelrod, K.C., Zaveri, J., Algire, M.A., Moodie, M., Montague, M.G., Venter, J.C., Smith, H.O., and Hutchison, C.A., One-step assembly in yeast of 25 overlapping DNA fragments to form a complete synthetic *Mycoplasma genitalium* genome. *Proceedings of the National Academy of Sciences*, *2008*. ***105****(51): p. 20404-20409*.
6. Beaucage, S. and Caruthers, M., Deoxynucleoside phosphoramidites—a new class of key intermediates for deoxypolynucleotide synthesis. *Tetrahedron letters*, *1981*. ***22****(20): p. 1859-1862*.
7. Hughes, R.A. and Ellington, A.D., Synthetic DNA synthesis and assembly: Putting the synthetic in synthetic biology. *Cold Spring Harbor Perspectives in Biology*, *2017*. ***9****(1): p. a023812*.
8. Lee, H., Wiegand, D.J., Griswold, K., Punthambaker, S., Chun, H., Kohman, R.E., and Church, G.M., Photon-directed multiplexed enzymatic DNA synthesis for molecular digital data storage. *Nature Communications*, *2020*. ***11****(1): p. 1-9*.
9. Lee, H.H., Kalhor, R., Goela, N., Bolot, J., and Church, G.M., Terminator-free template-independent enzymatic DNA synthesis for digital information storage. *Nature Communications*, *2019*. ***10****(1): p. 1-12*.
10. Ostrov, N., Beal, J., Ellis, T., Gordon, D.B., Karas, B.J., Lee, H.H., Lenaghan, S.C., Schloss, J.A., Stracquadanio, G., and Trefzer, A., Technological challenges and milestones for writing genomes. *Science*, *2019*. ***366****(6463): p. 310-312*.
11. Paul, S.S., Trabelsi, H., Yaseen, Y., Basu, U., Altaii, H.A., and Dhali, D., Advances in long DNA synthesis, in Microbial Cell Factories Engineering for Production of Biomolecules. 2021, *Elsevier*. p. 21-36.
12. Hao, M., Qiao, J., and Qi, H., Current and emerging methods for the synthesis of single-stranded DNA. *Genes*, *2020*. ***11****(2): p. 116*.
13. Eisenstein, M., Enzymatic DNA synthesis enters new phase. *Nature Biotechnology*, *2020*. ***38****(10): p. 1113-1115*.
14. Forster, A.C. and Church, G.M., Towards synthesis of a minimal cell. *Molecular Systems Biology*, *2006*. ***2****: p. 45*.
15. Chao, R., Yuan, Y., and Zhao, H., Recent advances in DNA assembly technologies. *FEMS Yeast Research*, *2014*
16. Casini, A., Storch, M., Baldwin, G.S., and Ellis, T., Bricks and blueprints: methods and standards for DNA assembly. *Nature Reviews Molecular Cell Biology*, *2015*. ***16****(9): p. 568-576*.
17. Gibson, D.G., Young, L., Chuang, R.-Y., Venter, J.C., Hutchison, C.A., and Smith, H.O., Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nature Methods*, *2009*. ***6****(5): p. 343-345*.
18. Gibson, D.G., Smith, H.O., Hutchison, C.A., Venter, J.C., and Merryman, C., Chemical synthesis of the mouse mitochondrial genome. *Nature methods*, *2010*. ***7****(11): p. 901-903*.
19. Karas, B.J., Suzuki, Y., and Weyman, P.D., Strategies for cloning and manipulating natural and synthetic chromosomes. *Chromosome Research*, *2015*. ***23****(1): p. 57-68*.
20. Libicher, K., Hornberger, R., Heymann, M., and Mutschler, H., *In vitro* self-replication and multicistronic expression of large synthetic genomes. *Nature communications*, *2020*. ***11****(1): p. 1-8*.
21. Van Nies, P., Westerlaken, I., Blanken, D., Salas, M., Mencia, M., and Danelon, C., Self-replication of DNA by its encoded proteins in liposome-based synthetic cells. *Nature Communications*, *2018*. ***9****(1): p. 1583*.
22. Mukai, T., Yoneji, T., Yamada, K., Fujita, H., Nara, S., and Su'etsugu, M., Overcoming the challenges of megabase-sized plasmid construction in *Escherichia coli*. *ACS Synthetic Biology*, *2020*. ***9****(6): p. 1315-1327*.
23. Su'etsugu, M., Takada, H., Katayama, T., and Tsujimoto, H., Exponential propagation of large circular DNA by reconstitution of a chromosome-replication cycle. *Nucleic Acids Research*, *2017*. ***45****(20): p. 11525-11534*.
24. Shao, Y., Lu, N., Wu, Z., Cai, C., Wang, S., Zhang, L.-L., Zhou, F., Xiao, S., Liu, L., and Zeng, X., Creating a functional single-chromosome yeast. *Nature*, *2018*. ***560****(7718): p. 331-335*.
25. Burke, D.T., Carle, G.F., and Olson, M.V., Cloning of large segments of exogenous DNA into yeast by means of artificial chromosome vectors. *Science*, *1987*. ***236****(4803): p. 806-812*.
26. Larionov, V., Kouprina, N., Graves, J., Chen, X., Korenberg, J.R., and Resnick, M.A., Specific cloning of human DNA as yeast artificial chromosomes by transformation-associated recombination. *Proceedings of the National Academy of Sciences*, *1996*. ***93****(1): p. 491-496*.
27. Kunes, S., Botstein, D., and Fox, M.S., Transformation of yeast with linearized plasmid DNA: formation of inverted dimers and recombinant plasmid products. *Journal of molecular biology*, *1985*. ***184****(3): p. 375-387*.
28. Noskov, V., Koriabine, M., Solomon, G., Randolph, M., Barrett, J., Leem, S.-H., Stubbs, L., Kouprina, N., and Larionov, V., Defining the minimal length of sequence homology required for selective gene isolation by TAR cloning. *Nucleic acids research*, *2001*. ***29****(6): p. e32-e32*.
29. Larionov, V., Kouprina, N., Eldarov, M., Perkins, E., Porter, G., and Resnick, M.A., Transformation-associated recombination between diverged and homologous DNA repeats is induced by strand breaks. *Yeast*, *1994*. ***10****(1): p. 93-104*.
30. Venetz, J.E., Del Medico, L., Wölfle, A., Schächle, P., Bucher, Y., Appert, D., Tschan, F., Flores-Tinoco, C.E., Van Kooten, M., and Guennoun, R., Chemical synthesis rewriting of a bacterial genome to achieve design flexibility and biological functionality. *Proceedings of the National Academy of Sciences*, *2019*. ***116****(16): p. 8070-8079*.
31. Fredens, J., Wang, K., De La Torre, D., Funke, L.F., Robertson, W.E., Christova, Y., Chia, T., Schmied, W.H., Dunkelmann, D.L., and Beránek, V., Total synthesis of *Escherichia coli* with a recoded genome. *Nature*, *2019*. ***569****(7757): p. 514-518*.
32. Thao, T.T.N., Labroussaa, F., Ebert, N., V'kovski, P., Stalder, H., Portmann, J., Kelly, J., Steiner, S., Holwerda, M., and Kratzel, A., Rapid reconstruction of SARS-CoV-2 using a synthetic genomics platform. *Nature*, *2020*. ***582****(7813): p. 561-565*.
33. Vashee, S., Arfi, Y., and Lartigue, C., Budding yeast as a factory to engineer partial and complete microbial genomes. *Current Opinion in Systems Biology*, *2020*. ***24****: p. 1-8*.
34. Benders, G.A., Noskov, V.N., Denisova, E.A., Lartigue, C., Gibson, D.G., Assad-Garcia, N., Chuang, R.-Y., Carrera, W., Moodie, M., and Algire, M.A., Cloning whole bacterial genomes in yeast. *Nucleic Acids Research*, *2010*. ***38****(8): p. 2558-2569*.
35. Mitchell, L.A., Mcculloch, L.H., Pinglay, S., Berger, H., Bosco, N., Brosh, R., Bulajić, M., Huang, E., Hogan, M.S., and Martin, J.A., *De novo* assembly and delivery to mouse cells of a 101 kb functional human gene. *Genetics*, *2021*. ***218****(1): p. iyab038*.
36. Pretorius, I. and Boeke, J., Yeast 2.0—connecting the dots in the construction of the world's first functional synthetic eukaryotic genome. *FEMS yeast research*, *2018*. ***18****(4): p. foy032*.
37. Eisenstein, M., How to build a genome. *Nature*, *2020*. ***578****(7796): p. 633-635*.
38. Dai, J., Boeke, J.D., Luo, Z., Jiang, S., and Cai, Y., Sc3.0:

revamping and minimizing the yeast genome. *Genome Biology*, *2020*. ***21****(1)*.

39. Sorek, R., Zhu, Y., Creevey, C.J., Francino, M.P., Bork, P., and Rubin, E.M., Genome-wide experimental determination of barriers to horizontal gene transfer. *Science*, *2007*. ***318****(5855): p. 1449-1452*.
40. Tagwerker, C., Dupont, C.L., Karas, B.J., Ma, L., Chuang, R.-Y., Benders, G.A., Ramon, A., Novotny, M., Montague, M.G., and Venepally, P., Sequence analysis of a complete 1.66 Mb *Prochlorococcus marinus* MED4 genome cloned in yeast. *Nucleic acids research*, *2012*. ***40****(20): p. 10375-10383*.
41. Itaya, M., Sato, M., Hasegawa, M., Kono, N., Tomita, M., and Kaneko, S., Far rapid synthesis of giant DNA in the *Bacillus subtilis* genome by a conjugation transfer system. *Scientific reports*, *2018*. ***8****(1): p. 8792*.
42. Postma, E.D., Dashko, S., Van Breemen, L., Taylor Parkins, S.K., Van Den Broek, M., Daran, J.-M., and Daran-Lapujade, P., A supernumerary designer chromosome for modular *in vivo* pathway assembly in *Saccharomyces cerevisiae*. *Nucleic Acids Research*, *2021*. ***49****(3): p. 1769-1783*.
43. Postma, E.D., Hassing, E.-J., Mangkusaputra, V., Geelhoed, J., De La Torre, P., Van Den Broek, M., Mooiman, C., Pabst, M., Daran, J.-M., and Daran-Lapujade, P., Modular, synthetic chromosomes as new tools for large scale engineering of metabolism. *Metabolic Engineering*, *2022*. ***72****: p. 1-13*.
44. Blount, B.A., Driessen, M.R.M., and Ellis, T., GC preps: Fast and easy extraction of stable yeast genomic DNA. *Scientific Reports*, *2016*. ***6****(1): p. 26863*.
45. Lee, K., Ji, J.-H., Yoon, K., Che, J., Seol, J.-H., Lee, S.E., and Shim, E.Y., Microhomology selection for microhomology mediated end joining in *Saccharomyces cerevisiae*. *Genes*, *2019*. ***10****(4): p. 284*.
46. Ranjha, L., Howard, S.M., and Cejka, P., Main steps in DNA double-strand break repair: an introduction to homologous recombination and related processes. *Chromosoma*, *2018*. ***127****(2): p. 187-214*.
47. Albarracín, V.H., Pathak, G.P., Douki, T., Cadet, J., Borsarelli, C.D., Gärtner, W., and Farias, M.E., Extremophilic acinetobacter strains from high-altitude lakes in argentinean puna: Remarkable UV-B resistance and efficient DNA damage repair. *Origins of Life and Evolution of Biospheres*, *2012*. ***42****(2-3): p. 201-221*.
48. Sato, T., Takada, D., Itoh, T., Ohkuma, M., and Atomi, H., Integration of large heterologous DNA fragments into the genome of *Thermococcus kodakarensis*. *Extremophiles*, *2020*. ***24****(3): p. 339-353*.
49. Dean, F.B., Rapid amplification of plasmid and phage DNA using phi29 DNA polymerase and multiply-primed rolling circle amplification. *Genome Research*, *2001*. ***11****(6): p. 1095-1099*.
50. Lau, Y.H., Stirling, F., Kuo, J., Karrenbelt, M.A., Chan, Y.A., Riesselman, A., Horton, C.A., Schäfer, E., Lips, D., and Weinstock, M.T., Large-scale recoding of a bacterial genome by iterative recombineering of synthetic DNA. *Nucleic acids research*, *2017*. ***45****(11): p. 6971-6980*.
51. Kwon, Y., Daley, J.M., and Sung, P., Reconstituted system for the examination of repair DNA synthesis in homologous recombination. *Methods in enzymology*, *2017*. ***591****: p. 307-325*.
52. Shimizu, Y., Inoue, A., Tomari, Y., Suzuki, T., Yokogawa, T., Nishikawa, K., and Ueda, T., Cell-free translation reconstituted with purified components. *Nature biotechnology*, *2001*. ***19****(8): p. 751*.
53. Guo, Z.-P., Borsenberger, V., Croux, C., Duquesne, S., Truan, G., Marty, A., and Bordes, F., An artificial chromosome yIAC enables efficient assembly of multiple genes in *Yarrowia lipolytica* for biomanufacturing. *Communications biology*, *2020*. ***3****(1): p. 1-10*.
54. Chen, W., Han, M., Zhou, J., Ge, Q., Wang, P., Zhang, X., Zhu, S., Song, L., and Yuan, Y., An artificial chromosome for data storage. *National Science Review*, *2021*. ***8****(5): p. nwab028*.
55. Boeke, J.D., Church, G., Hessel, A., Kelley, N.J., Arkin, A., Cai, Y., Carlson, R., Chakravarti, A., Cornish, V.W., Holt, L., Isaacs, F.J., Kuiken, T., Lajoie, M., Lessor, T., Lunshof, J., Maurano, M.T., Mitchell, L.A., Rine, J., Rosser, S., Sanjana, N.E., Silver, P.A., Valle, D., Wang, H., Way, J.C., and Yang, L., The Genome Project-Write. *Science*, *2016*. ***353****(6295): p. 126-127*.
56. Mutschler, H., Robinson, T., Tang, T.Y.D., and Wegner, S., Special issue on bottom-up synthetic biology. *ChemBioChem*, *2019*. ***20****(20): p. 2533-2534*.
57. Belda, I., Williams, T.C., De Celis, M., Paulsen, I.T., and Pretorius, I.S., Seeding the idea of encapsulating a representative synthetic metagenome in a single yeast cell. *Nature Communications*, *2021*. ***12****(1): p. 1-8*.
58. Oldfield, L.M., Grzesik, P., Voorhies, A.A., Alperovich, N., Macmath, D., Najera, C.D., Chandra, D.S., Prasad, S., Noskov, V.N., Montague, M.G., Friedman, R.M., Desai, P.J., and Vashee, S., Genome-wide engineering of an infectious clone of herpes simplex virus type 1 using synthetic genomics assembly methods. *Proceedings of the National Academy of Sciences*, *2017*. ***114****(42): p. E8885-E8894*.
59. Shang, Y., Wang, M., Xiao, G., Wang, X., Hou, D., Pan, K., Liu, S., Li, J., Wang, J., and Arif, B.M., Construction and rescue of a functional synthetic baculovirus. *ACS synthetic biology*, *2017*. ***6****(7): p. 1393-1402*.
60. Vashee, S., Stockwell, T.B., Alperovich, N., Denisova, E.A., Gibson, D.G., Cady, K.C., Miller, K., Kannan, K., Malouli, D., and Crawford, L.B., Cloning, assembly, and modification of the primary human cytomegalovirus isolate Toledo by yeast-based transformation-associated recombination. *Msphere*, *2017*. ***2****(5): p. e00331-17*.
61. Gibson, D.G., Glass, J.I., Lartigue, C., Noskov, V.N., Chuang, R.-Y., Algire, M.A., Benders, G.A., Montague, M.G., Ma, L., and Moodie, M.M., Creation of a bacterial cell controlled by a chemically synthesized genome. *Science*, *2010*. ***329****(5987): p. 52-56*.
62. Ruiz, E., Talenton, V., Dubrana, M.-P., Guesdon, G., Lluch-Senar, M., Salin, F., Sirand-Pugnet, P., Arfi, Y., and Lartigue, C., CReasPy-cloning: a method for simultaneous cloning and engineering of megabase-sized genomes in yeast using the CRISPR-Cas9 system. *ACS synthetic biology*, *2019*. ***8****(11): p. 2547-2557*.
63. Rideau, F., Le Roy, C., Descamps, E.C., Renaudin, H.L.N., Lartigue, C., and BéBéAr, C.C., Cloning, stability, and modification of *Mycoplasma hominis* genome in yeast. *ACS synthetic biology*, *2017*. ***6****(5): p. 891-901*.
64. Karas, B.J., Tagwerker, C., Yonemoto, I.T., Hutchison Iii, C.A., and Smith, H.O., Cloning the *Acholeplasma laidlawii* PG-8A genome in *Saccharomyces cerevisiae* as a yeast centromeric plasmid. *ACS synthetic biology*, *2012*. ***1****(1): p. 22-28*.
65. Zhou, J., Wu, R., Xue, X., and Qin, Z., CasHRA (Cas 9-facilitated Homologous Recombination Assembly) method of constructing megabase-sized DNA. *Nucleic acids research*, *2016*. ***44****(14): p. e124-e124*.
66. Noskov, V.N., Karas, B.J., Young, L., Chuang, R.-Y., Gibson, D.G., Lin, Y.-C., Stam, J., Yonemoto, I.T., Suzuki, Y., and Andrews-Pfannkoch, C., Assembly of large, high G+C bacterial DNA fragments in yeast. *ACS synthetic biology*, *2012*. ***1****(7): p. 267-273*.
67. Karas, B.J., Molparia, B., Jablanovic, J., Hermann, W.J., Lin, Y.-C., Dupont, C.L., Tagwerker, C., Yonemoto, I.T., Noskov, V.N., and Chuang, R.-Y., Assembly of eukaryotic algal chromosomes in yeast. *Journal of biological engineering*, *2013*. ***7****(1): p. 30*.
68. O'neill, B.M., Mikkelson, K.L., Gutierrez, N.M., Cunningham, J.L., Wolff, K.L., Szyjka, S.J., Yohn, C.B., Redding, K.E., and Mendez, M.J., An exogenous chloroplast genome for complex sequence manipulation in algae. *Nucleic acids research*, *2012*. ***40****(6): p. 2782-2792*.
69. Zhang, W., Zhao, G., Luo, Z., Lin, Y., Wang, L., Guo, Y., Wang, A., Jiang, S., Jiang, Q., and Gong, J., Engineering the ribosomal DNA in a megabase synthetic chromosome. *Science*, *2017*. ***355****(6329)*.



CHAPTER 3

LONG-READ DIRECT RNA SEQUENCING OF THE MITOCHONDRIAL TRANSCRIPTOME OF *SACCHAROMYCES CEREVISIAE* REVEALS CONDITION-DEPENDENT INTRON ABUNDANCE

Charlotte C. Koster, Askar Kleefeldt, Marcel van den Broek, Marijke Luttik, Jean-Marc Daran & Pascale Daran-Lapujade

This chapter was deposited as a preprint at BioRxiv
doi: https://doi.org/10.1101/2023.01.19.524680

Abstract

Mitochondria fulfil many essential roles and have their own genome, which is expressed as polycistronic transcripts that undergo co- or post-transcriptional processing and splicing. Due to inherent complexity and limited technical accessibility of the mitochondrial transcriptome, fundamental questions regarding mitochondrial gene expression and splicing remain unresolved, even in the model eukaryote *Saccharomyces cerevisiae*. Long-read sequencing could address these fundamental questions. Therefore, a method for enrichment of mitochondrial RNA and sequencing using Nanopore technology was developed, enabling the resolution of splicing of polycistronic genes and the quantification the spliced RNA.
This method successfully captured the full mitochondrial transcriptome and resolved RNA splicing patterns with single-base resolution, and was applied to explore the transcriptome of *S. cerevisiae* grown with glucose or ethanol as sole carbon source, revealing the impact of growth conditions on mitochondrial RNA-expression and splicing. This study uncovered a remarkable difference in turn-over of group II introns between yeast grown in mostly fermentative and fully respiratory conditions. Whether this accumulation of introns in glucose medium has an impact on mitochondrial functions remains to be explored. Combined with the high tractability of the model yeast *S. cerevisiae*, the developed method enables to explore mitochondrial transcriptome regulation and processing in a broad range of conditions relevant in human context, including aging, apoptosis and mitochondrial diseases.

3.1 Introduction

Mitochondria are one of the hallmarks of eukaryotic cells. Commonly known as 'the powerhouse of the cell', they perform several key cellular processes, such as the biosynthesis of iron-sulfur clusters, branched amino acids, heme, and lipids and play an important role in cellular aging and programmed cell death (reviewed by [1, 2]). To date, about 150 mitochondrial disorders have been reported in humans, which genetic origin lies in mitochondrial (mtDNA) or nuclear (nDNA) DNA. As a remnant of their bacterial origin, mitochondria harbor their own DNA, however most of the 1.000 – 1.500 proteins composing mitochondria are nuclear-encoded and imported from the cytosol to mitochondria [3]. Remarkably, mitochondria have retained specific genes and require their own complete set of proteins for gene expression, RNA splicing and degradation and DNA maintenance [4]. Mitochondrial functions are largely conserved across eukaryotes, including the eukaryotic model *Saccharomyces cerevisiae*, which is a preferred model for mitochondrial biology for two main reasons. Firstly, *S. cerevisiae* can grow in the complete absence of respiration and with partial or complete loss of mtDNA [5]. Secondly, *S. cerevisiae* is still to date one of the rare organisms harboring genetic tools to modify mtDNA beyond base editing (e.g., gene deletion, integration [6]). Using these features, mtDNA-free yeast cells (ρ^0 mutants) were shown to stably maintain mouse mtDNA [7], a first step towards the potential humanization and engineering of mtDNA in yeast. *S. cerevisiae* has therefore an important role to play in deepening our understanding of mitochondrial processes.

The *circa* 86 kb *S. cerevisiae* mitochondrial genome encodes seven oxidative phosphorylation proteins (*COX1*, *COX2*, *COX3*, *COB*, *ATP6*, *ATP8*, *Q0255*), two ribosomal subunits (15S rRNA and 21S rRNA), a ribosomal protein (*VAR1*), the RNA subunit of RNase P (*RPM1*), and a full set of 24 tRNAs [8]. Unlike its nDNA, *S. cerevisiae* mtDNA has a prokaryote-like structure with 11 polycistronic primary transcripts (Figure 3.1), which expression is regulated by T7-like promoters [9, 10]. Conserved dodecamer sequences separate the transcripts within the polycistrons and enable modulation of gene expression, while transcription is initiated at conserved nonanucleotide promoters [11-13]. While splicing is rare in *S. cerevisiae* nDNA [14], three mitochondrial genes (*COX1*, *COB* and *21S rRNA*) contain intronic sequences [8, 15]. *Q0255* requires the omission of two GC-clusters to be fully translated [16, 17]. Several of these introns encode proteins, e.g., the I-SceI homing endonuclease is encoded in the intron of the 21S rRNA gene [18], while introns of *COX1* and *COB* encode maturases that assist in splicing of their respective intron sequences [19]. Some intron-encoded proteins are the result of alternative splicing, where parts of an exon combined with different intronic sequences of the same primary transcript yield different proteins, next to the exon-encoded protein.

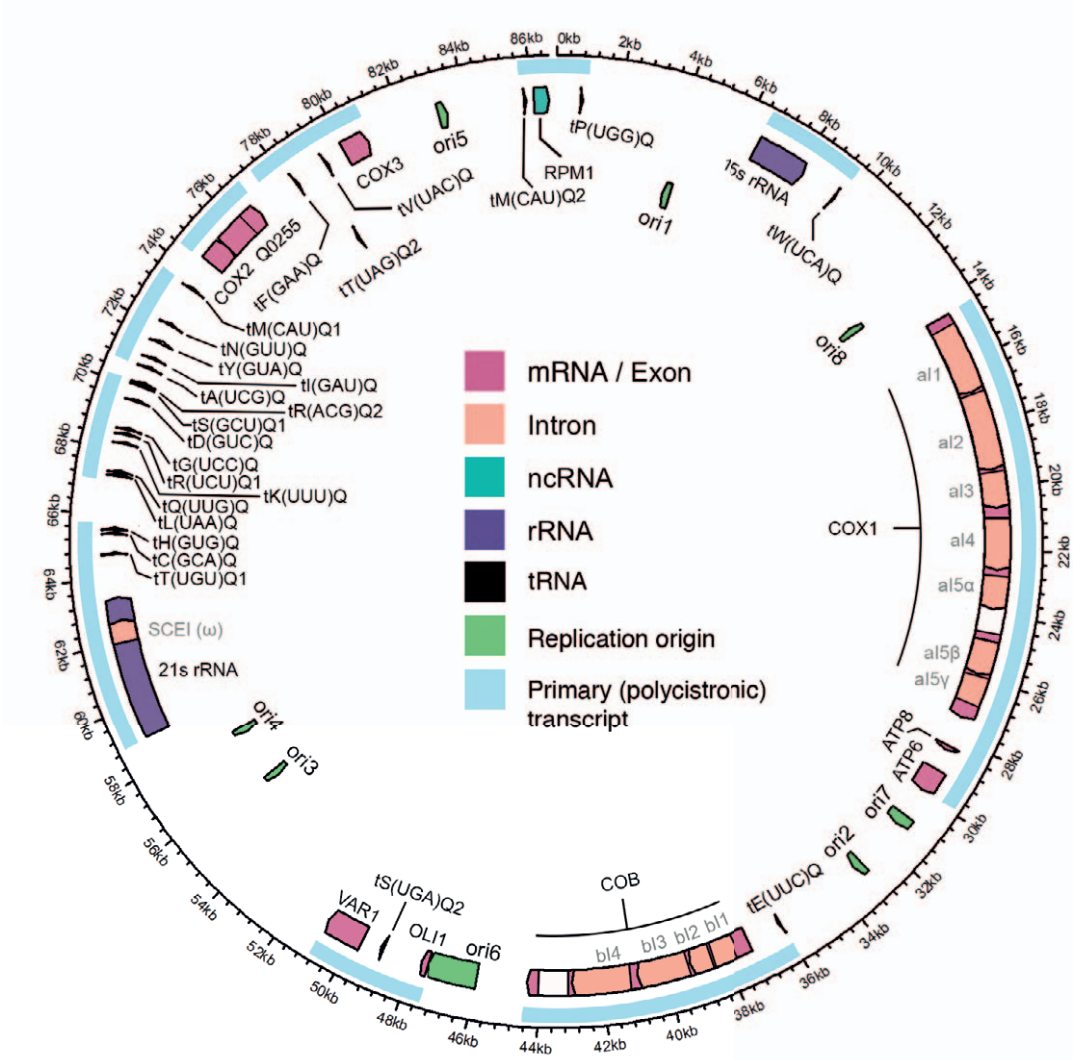


Figure 3.1. Annotated mitochondrial genome of CEN.PK113-7D. Primary polycistronic transcripts (light blue boxes); mRNA: messenger RNA, ncRNA: noncoding RNA, rRNA: ribosomal RNA, tRNA: transfer RNA. Exons that encode the main proteins are indicated in pink. Coding intron sequences are encoded in orange, non-coding intron sequences are indicated in white.

These splicing events are catalyzed by nDNA- and mtDNA-encoded enzymes [20]. Despite its small size, the mtDNA results in a complex RNA landscape involving mechanisms that, despite decades of study, are not fully elucidated.

Since the first gene arrays in the '90s, the cytosolic transcriptome of *S. cerevisiae* has been intensively investigated [21, 22]. The mitochondrial transcriptome is generally not detectable in these classical transcriptome studies, which is largely explained by the fact that mRNA extraction methods rely on selection of poly(A)-tailed mRNA, while yeast mtRNA is not polyadenylated [23]. Additionally, the mitochondrial transcriptome represents a small fraction of the total yeast transcriptome (ca. 5 %), which is largely dominated by (ribosomal) cytosolic RNA species [24, 25]. Capturing the mitochondrial transcriptome therefore requires a dedicated methodology. To date, a single study reported the yeast mitochondrial transcriptome, using cellular subfractionation to enrich for mitochondria and short read RNA sequencing [13]. While this study brought new insight in regulatory elements such as promoter sequences, dodecamer sequences, UTRs and

processing sites, the use of short reads, even with paired ends, cannot capture the full complexity of the mtRNA transcriptome. This technical limitation can be overcome by implementing long-read RNA sequencing, as recently demonstrated by the successful resolution of various complex and spliced transcriptomes ranging from viruses to plant, mammalian and even a full poly(A)-human transcriptome [22, 26-31].

The goal of this study is to provide a comprehensive description of the mitochondrial transcriptome of *Saccharomyces cerevisiae*, including identification of processing of polycistronic transcripts and splicing of the different transcript isoforms. To this end, we developed a robust protocol for mitochondrial RNA isolation, and combined it with Nanopore long-read direct RNA sequencing technology. This method was used to investigate the response of the mitochondrial transcriptome to different carbon sources (glucose and ethanol) chosen for their ability to tune yeast physiology and respiratory activity.

3.2 Materials & Methods

Strains & culture conditions

Saccharomyces cerevisiae strains used in this study (Table 3.1, Table S3.1) were derived from the CEN.PK lineage [32], or from a 161 lineage (also known as ID41-6/161) [33-35], which were a gift from prof. Alan Lambowitz (University of Texas at Austin, Austin, TX, USA).
Yeast strains were grown aerobically at 30 °C in 500 mL shake flasks containing 100 mL synthetic medium with ammonium as nitrogen source (SM) supplied with vitamins and trace elements, prepared and sterilized as described previously [36], in Innova incubator shakers (Eppendorf, Hamburg, Germany) set at 200 rpm. For respiro-fermentative growth sterilized media were supplemented with a D-glucose solution to a final concentration of 20 g L^{-1} glucose (SMD, 'glucose media'). For respiratory growth, media were supplemented with absolute ethanol to a final concentration of 2 % (v/v) and with a glycerol solution to a final concentration of 2 % (v/v) (SMEG, 'ethanol media'). Glucose solutions (50 % w/v) and glycerol solutions (99 % w/v) were autoclaved separately for 10 minutes at 110 °C. When required, medium was supplemented with separately sterilized solutions of uracil (Ura) to a final concentration of 150 mg L^{-1}, Lysine (Lys) to a final concentration of 300 mg L^{-1} and/or Adenine (Ade) to a final concentration of 200 mg L^{-1} [37]. Alternatively, YPD medium, containing 10 g L^{-1} Bacto Yeast extract, 20 g L^{-1} Bacto Peptone and 20 g L^{-1} glucose was used.
For plate cultivation 2 % (w/v) agar was added to the medium prior to heat sterilization. Frozen stocks were prepared by addition of sterile glycerol (30 % v/v) to late exponential phase shake-flask cultures of *S. cerevisiae* and 1 mL aliquots were stored aseptically at –80 °C.

Table 3.1. *S. cerevisiae* strains used in this study.

Strain name	*Relevant genotype*	*Parental strain*	*Reference*
CEN.PK113-7D	*MATa*	-	Entian and Kötter [32]
CEN.PK113-5D	*MATa ura3-52*	-	Entian and Kötter [32]
161-U7 / 1^+2^+	*MATa ade1 lys1 ura3*, no mitochondrial ω-intron	-	Mahler and Lin [33]
161-U7 I^0	*MATa ade1 lys1 ura3*, no mitochondrial ω-intron, no introns in *COX1* and *COB*	161-U7 / 1^+2^+	Huang *et al.* [35]
IMC251	*MATa ura3-52* + pUDC329 (*preCOX4-mTq2, URA3*)	CEN.PK113-5D	This study
IMC252	*MATa ade1 lys1 ura3*, no mitochondrial w-intron + pUDC329 (*preCOX4-ymTq2, URA3*)	161-U7 / 1^+2^+	This study
IMC253	*MATa ade1 lys1 ura3*, no mitochondrial w-intron, no introns in *COX1* and *COB* + pUDC329 (*preCOX4-ymTq2, URA3*)	161-U7 I^0	This study

Molecular biology & strain construction

Strains expressing the yeast-optimized fluorescent protein ymTurquoise2 (*ymTq2*, [38]) in the mitochondria (IMC251, IMC252 and IMC253, Table 3.1) were transformed with plasmid pUDC329 (*pTEF1-preSU9-ymTq2-tENO2*, *URA3*, *CEN/ARS*, *bla*, *ColE1)*. Plasmid pUDC329 was cloned using Golden-Gate assembly performed according to Lee *et al.* [39], from part plasmids pUD538 (GFP-dropout backbone, *URA3*,

CEN/ARS, *bla*, *ori)*, pYTK013 (part 2, *pTEF1*), pGGkp320 (part 3a, *preSU9*), pGGkp308 (part 3b, *ymTurquoise2*), pYTK055 (part 4, *tENO2*) (Table S3.2) [40, 41]. The part plasmids were either obtained from the Yeast Toolkit collection [39] or made according to the author's instructions using primers with Golden Gate flanking regions corresponding to the respective part type. The *preSU9* sequence was ordered as a synthetic sequence at GeneArt (Supplementary Information, Life technologies, Carlsbad, CA). pDRF1-GW ymTurquoise2, *pFA6a-link-ymNeongreen-URA3* and pDRF1-GW ymYPET were a gift from Bas Teusink (Addgene plasmid #118453, #125703, #118455; [38]). 1 µL of the Golden-Gate reaction mixture was transformed into *E. coli* cells (XL1-Blue, Agilent Technologies, Santa Clara, CA), which were grown in solid Lysogeny broth (LB) medium (5.0 g L^{-1} yeast extract, 10 g L^{-1} Bacto tryptone, 2 % Bacto agar (BD Biosciences, Franklin Lakes, NJ), 5.0 g L^{-1} NaCl) supplemented with 100 mg L^{-1} ampicillin. Upon growth, random colonies were picked and resuspended in 10 µL sterile water, of which 1 µL was used for diagnostic colony PCR, using using DreamTaq MasterMix 2X (Thermo Fisher Scientific, Waltham, MA)) and desalted primers (Sigma-Aldrich, St. Louis, MO, Table S3.3), with an initial 10-minute incubation at 95 °C to release *E. coli* DNA. Positive transformants were inoculated in 5 mL liquid LB medium (omitting the Bacto agar) and plasmids were isolated using a GeneJET Plasmid Miniprep Kit (Thermo Fisher Scientific). Correct plasmid structure was confirmed using restriction analysis using the FastDigest KpnI enzyme (Thermo Fisher Scientific) according to the manufacturer's instructions and by EZseq Sanger sequencing (Macrogen, Amsterdam, the Netherlands). Fragment sizes were assessed on a 1 % agarose gel. 100 ng plasmid was transformed into *S. cerevisiae* cells as previously described [42], transformants were selected on SMD (+Lys/Ade) medium lacking uracil. Successful transformation with the plasmid was confirmed using fluorescence microscopy, transformants were re-streaked three times on selective solid medium to obtain single colony isolates, which were grown and stored as described before. Construction of strain IMC173, used for optimization of the mitochondrial isolation protocol (Figure S3.1, Table S3.1) is described in the Supplementary Information [43, 44].

Preparation of mitochondrial RNA & sequencing

Isolation of mitochondria by differential centrifugation

Crude mitochondria were isolated by a combination of protocols described by [45] and [46], with a few modifications. The protocol was initially tested using a strain (IMC173) with fluorescent mitochondria prior to isolating mitochondria for RNA-seq (Figure S3.1, Supplementary methods). Pre-cultures of CEN.PK113-7D were grown overnight on SMD and were used to inoculate duplicate 500 mL shake flasks containing 100 mL SMD or SMEG at an optical density at 660 nm (OD_{660}) of 0.1 - 0.2. The mitochondrial volume of a cell is dependent on respiratory activity [47], therefore, to yield enough mitochondrial mass, 300 mL of culture was used when grown on respiratory SMEG media and 800 mL of culture for respiro-fermentative glucose media. SMD cultures were grown until mid-exponential phase (OD_{660} of 5) to prevent cells from going through diauxic shift. SMEG cultures were grown until mid- to late-exponential phase (OD_{660} of 5 - 19), before harvesting by centrifugation. All centrifugation steps were performed at 4 °C in an Avanti J-E high-speed fixed-angle centrifuge (Beckman Coulter Life Sciences, Indianapolis, IN) with a JA-10 rotor for volumes larger than 30 mL and a JA-25.50 rotor for 30 mL and less. Stock solutions were used to prepare working solutions and were sterilized prior to use at 121 °C for 30 minutes. The following stock solutions were prepared: potassium phosphate buffer pH 7.4 (30.4 mM KH_2PO_4, 69.6 mM K_2HPO_4), Tris buffer (0.1 M Tris, pH 7.4), HEPES-KOH (0.5 M HEPES, pH set to 7.2 with KOH), $MgCl_2 \cdot 6H_2O$ (0.1 M). A sorbitol stock solution (3 M) was autoclaved at 110 C for 10 minutes. Working solutions were freshly prepared prior to use.

Cultures were centrifuged in pre-weighed bottles for 5 min at 1,500 × *g* to remove media, then washed with Tris (0.1 M Tris, pH 7.4) and centrifuged again. The supernatant was thoroughly removed, the weight of each pellet was determined, and pellets were resuspended in 30 mL Tris-DTT (0.1 M Tris, 10 mM 1,4-Dithiothreitol (DTT), prepared fresh before use, pH 9.3) and transferred to centrifuge tubes. Cultures were incubated in Tris-DTT for 15 minutes in a waterbath at 30 °C with gentle shaking, then centrifuged 10 min at 4,342 × *g*, washed once in Sorbitol-Phosphate buffer (SP-buffer, 1.2 M sorbitol, 20 mM potassium phosphate buffer pH 7.4) and resuspended in SP-buffer. Zymolyase 20T (AMSbio, Alkmaar, the Netherlands) was added to each culture at a concentration of 5 – 10 mg per gram pellet wet weight. Cultures were incubated in presence of Zymolyase in a waterbath at 30 °C with gentle shaking to obtain spheroplasts. Spheroplasting progress

was monitored by resuspending 10 µL of spheroplasts in 2 mL of dH_2O and measuring the OD_{660}, as the low osmotic pressure of water will cause bursting of spheroplasts. Spheroplasting was halted after the OD_{660} dropped by 50 - 75 %, (ca. 30 – 45 min). Spheroplasts were always kept on ice.
Spheroplasts were pelleted 10 min at 4,342 × *g*, and supernatant containing Zymolyase was thoroughly removed. Spheroplasts were then resuspended in 20 mL SP-buffer by gentle shaking and pelleted again (10 min at 4,342 × *g*). The washed and pelleted spheroplasts were resuspended in low-osmolarity Sorbitol-HEPES buffer (SHE, 0.6 M sorbitol, 20 mM HEPES-KOH, 2 mM $MgCl_2$) with with cOmplete Protease Inhibitors Cocktail (PI, Roche Diagnostics, Rotkreuz, Switzerland) and homogenized at 4 °C using a pre-cooled Potter-Elvehjem PTFE pestle at 100 RPM and glass tube with a working volume of 30 mL. The homogenate was cleared of cellular debris by centrifugation at 1,500 × *g* for 10 min. The supernatant was collected with a pipette, transferred to a clean tube, and centrifuged at 12,000 × *g*. The supernatant containing the cytosolic fraction was collected in a separate tube. The pellet, containing the mitochondria-enriched fraction was resuspended in 2 mL of SHE + PI and immediately processed for RNA extraction or flash-frozen in liquid nitrogen and stored at -80 °C.

RNA extraction from mitochondrial fraction

RNA-extraction and handling was done in an RNase-free workspace. When possible, materials and workspaces were cleaned with RNaseZAP spray and towels (Sigma Aldrich, St. Louis, MO) and consumables and working solutions were autoclaved for 45 minutes at 121 °C to denature any RNases. *In vitro* synthesized RNA was spiked throughout the protocol to assess throughput and efficiency of the protocol. The synthesis of control RNA and timing of spike-ins is described in the supplementary methods. RNA from the mitochondria-enriched fraction was isolated using the miRNeasy Mini kit (Qiagen, Venlo, the Netherlands). Mitochondrial fractions were spun down for 10 minutes 10,000 × *g* at 4 °C, the buffer was removed, and the mitochondria were thoroughly resuspended in 700 µL QIAzol reagent at RT by vortexing. This was sufficient to lyse the mitochondria, and RNA was extracted following manufacturer's instruction. On-column DNase I treatment using the RNase-free DNase set (Qiagen) was included in the RNA extraction to remove any DNA contaminants. RNA was eluted in 30 µL nuclease-free water at RT. RNA integrity (RIN) was assessed using an RNA ScreenTape assay for the Tapestation (Agilent, Santa Clara, CA), and RNA quantity using a Qubit fluorometer with the RNA broad range assay kit (Thermo Fisher, Waltham, MA). Samples with an RIN > 7 and a yield of 3 µg or more were used for enzymatic polyadenylation.

In vitro polyadenylation and cleanup of isolated RNA

In vitro polyadenylation was performed using *E. coli* Poly(A) Polymerase (New England Biolabs, Ipswich, MA), according to manufacturer's instructions, with a few modifications: 3 – 10 µg of RNA was added to each reaction to meet Nanopore input requirements. Additionally, 20 U murine RNase inhibitor (New England Biolabs) were added to each reaction. After polyadenylation, samples were cleaned up miRNeasy Mini kit with a few modifications. The polyadenylation reaction was filled up to 100 µL using nuclease-free water. Then 350 L RLT buffer (RNeasy kit, Qiagen) were added to the sample to replace QIAzol buffer. Next, 250 µL 100 % ethanol were added and the whole volume was transferred to a miRNeasy spin column. Column purification was further performed according to manufacturer's instructions, omitting the on-column DNase digestion. The RNA was eluted in 30 µL nuclease-free water and RIN was determined with an RNA Screentape assay, and quantity with a Qubit fluorometric assay. Presence of contaminants was determined by measuring absorbance using a Nanodrop spectrophotometer (Thermo Fisher). Purified RNA with an RIN of > 7, a concentration of > 55 ng/µL, a 260/280 absorbance ratio of around 2.0 and a 260/230 ratio between 2.0 and 2.2 was deemed high quality and used for RNA sequencing.

In vitro synthesis of RNA spike-in controls

RNA spike-in control sequences were generated by *in vitro* transcription (IVT) of DNA templates bearing a long T7 promoter. Three sequences were of 60 bp length (IVT01, IVT02, IVT03), one of 102 bp sequence (IVT04), one 729 bp sequence (IVT-Neon), one 742 bp sequence (IVT-YPet, [48]) and the 1026 bp IVT-eYFP sequence (IVT-YFP) (see: *"Synthetic RNA sequences used in this study"*, SI). DNA templates for 60 bp sequences (IVT01, IVT02 and IVT03) were generated by annealing of primers (Table S3.2) by heating

equimolar amounts of both primers up to 95 °C for 5 minutes and cooling down slowly at room temperature for primers to hybridize. DNA template for IVT04 was generated by overlap PCR, where self-dimerizing primers were used in final concentrations of 0.2 µM in PCR reactions of 50 µL reaction volume using Phusion High fidelity polymerase (Thermo Fisher) according to manufacturer's instructions. DNA templates for IVT-eYFP, IVT-Neon and IVT-YPet were generated by PCR amplification from plasmids using Phusion HF polymerase (Table S3.2, Table S3.4). All DNA templates were subject to PCR purification and size validated by agarose gel-electrophoreses. 0.5 – 1 µg DNA template were *in vitro* transcribed using HiScribe® T7 Quick High Yield RNA Synthesis Kit (New England Biolabs, Ipswich, MA), following manufacturer's instruction: IVT reactions with DNA templates < 300 bp were performed in reaction volume of 20 µL and reactions were incubated at 37 °C in a thermocycler for 4 – 16 hours. For DNA templates > 300 bp, the reaction volume was 30 µL, and incubation period was adjusted to 2 hours for 37 °C in a thermocycler. IVT products were recovered directly after reaction by miRNeasy Mini Kit for products < 300 bp, where reactions were filled to 100 µL with nuclease-free water and buffer RLT from RNeasy Mini Kit (Qiagen, Venlo, the Netherlands) was added, omitting QIAzol. Alternatively, IVT products < 300 bp were purified by lithium chloride (LiCl) precipitation as described in manufacturer's instructions. For IVT products > 300 bp, an RNeasy Mini Kit (Qiagen) was used, following manufacturer's instructions. All IVT products were eluted in 60 µL nuclease-free water. Quantity and purity were assessed photospectrometrically by Nanodrop and Qubit. IVT product sizes were validated by denaturing gel-electrophoresis with denaturing 2 % (w/v) agarose with 3 % (v/v) formaldehyde gels in MOPS buffer (40 mM MOPS pH 7, 10 mM sodium acetate, 1mM EDTA). Prior to loading on gel, 2X loading dye (NEB) was added the RNA and the sample was denatured at 70 °C for 10 minutes. Samples were directly put on ice for 2 minutes, and subsequently loaded onto the gel. 45 fmol of synthetic RNA controls were spiked in at two different timepoints: i) after isolation of the mitochondrial fraction, prior to RNA isolation, to simulate the presence of cytosolic RNA and ii) prior to polyadenylation, to assess the throughput and efficiency of polyadenylation, cleanup, and library preparation. Lastly, the ONT Direct RNA sequencing kit includes a synthetic *ENO2* control RNA, which was spiked during library preparation according to manufacturer's instructions.

Library preparation & sequencing.

Poly(A)-tailed RNA samples were used for library preparation using the Direct RNA Sequencing Kit (SQK-RNA002, Oxford Nanopore Technologies (ONT), Oxford, United Kingdom), following manufacturer's instructions. Library preparation yield was measured with Qubit® 2.0 fluorometer and corresponding Qubit™ dsDNA BR Assay Kit. Prior to library loading, remaining active pores were measured by a flow cell check with MinKNOW software (ONT). Library was loaded onto a R.9.4.1 flow cell with FLO-MIN106D chemistry using the Flow Cell Priming Kit (EXP-FLP002, ONT) according to manufacturer's instructions. Sequencing runs were started via MinKNOW with default parameters for 16 hours.

Computational methods for RNA sequencing data analysis

Availability of data

R scripts and CEN.PK113-7D mitochondrial reference sequences are deposited and can be accessed freely at https://gitlab.tudelft.nl/charlottekoste/mitornaseq. All raw sequencing data is deposited at NCBI (www.ncbi.nlm.nih.gov): RNA-seq data and normalized expression levels are deposited at GEO under accession number GSE219013 and DNA sequencing data as well as *de novo* assemblies are deposited under BioProject ID PRJNA902953. Supplementary file SI_readcounts.xlsx contains a spreadsheet with raw read counts per gene, gene counts normalized to CPM and gene counts normalized to CPM and mtDW, and is available through: https://doi.org/10.1101/2023.01.19.524680

Basecalling

Raw FAST5 files were basecalled on a high-performance computing cluster (GPU) using Guppy 4.4.1 (ONT) with default parameters for flow cell type and sequencing kit. Basecalled reads were categorized as PASSED for Guppy-determined quality score > 7 and as FAILED for reads < 7 (Table S3.5). All PASSED FASTQ files were concatenated into a single FASTQ files and filtered by length to cut-off at 40 bp basecalled reads were processed according to standard EPI2ME workflow (Metrichor Ltd., ONT, Oxford, UK) for FASTQ RNA

control reads to assess sequencing accuracy and coverage of the RNA control strand (RCS).

Annotation of the CEN.PK113-7D mitochondrial genome

As reference for alignment, *S. cerevisiae* CEN.PK113-7D complete genome sequence ([49, 50]) was used. The mitochondrial genome of CEN.PK113-7D was annotated by aligning nucleotide sequences of the S288C mitochondrial mapping data from the mitochondrial genome deposited in the *Saccharomyces* Genome Database (SGD: https://www.yeastgenome.org/ [51, 52]) as well as the S288C mitochondrial genome annotated by Turk *et al.* [13] to the CEN.PK113-7D mitochondrial genome sequence using SnapGene software (http://www.snapgene.com, GSL Biotech, San Diego, CA). The resulting annotated mitochondrial CEN.PK113-7D genome (Figure 3.1) was exported as GenBank and annotation (.gff) file and is deposited at Gitlab. The map of the CEN.PK113-7D mitochondrial genome was visualized in R using the circlize package [53].

Mapping

To account for control sequences, the FASTA file of the genome reference sequence was amended with FASTA files of all *in vitro* synthesized control sequences including the sequence of RCS (*ENO2* gene as provided by ONT). The resulting FASTA file and corresponding annotation file were used for mapping with minimap2 [54] (parameters: -ax splice -uf -k14 --secondary=no) to obtain a Sequence Alignment/Map (SAM) file [55]. All PASSED reads longer 40 bp were aligned to the reference FASTA file. SAM files were filtered and converted to Binary Alignment/Map (BAM) files and indexed using SAMtools, whereby only uniquely aligning reads were extracted based on FLAGs 0 and 16.
Poly(A)-tail length was estimated by nanopolish-polya [29], for both PASSED and FAILED reads of one of the sequencing runs (Ethanol replicate #3), following provided instructions. For assessment and visualization of poly(A)-tail lengths of datasets, only poly(A)-tail estimations with nanopolish-polya specific QC-tag PASS were used.

Feature quantification and normalization

Annotation of alignment and quantification of features was performed using FeatureCounts [56], where above named reference file was supplied together with filtered BAM file with uniquely aligned reads. Importantly, long-read specific, overlap allowed, directional counting in terms of strandedness and fractional counting (parameters: -L -O -s 1 --fraction) were employed. The present datasets originated from three biological culture replicates, mitochondria isolation rounds and sequencing runs, leading to different read distributions and library sizes (Figure 3.2B, Table S3.5). Quantitative analysis of expression levels therefore required library normalization. RPKM (Reads Per Kilobase Million) is the standard normalization method for RNA-seq, developed for short-read RNA sequencing in which read length (ca. 150 - 200 bp) is substantially shorter than gene length (e.g., several kb in yeast). RPKM normalizes to library size, but also to gene length, as to compensate for this direct correlation between read number and gene length. However, normalization by gene length is not applicable to long-read RNA sequencing as read length often approaches or even exceeds the gene length. The datasets were therefore only normalized over library depth in Counts Per Million (CPM) in R using edgeR [57]. For all quantification analysis, reads mapping to both the *S. cerevisiae* native *ENO2* feature as well as RCS feature, were considered to be control reads, as the two could not be distinguished based on sequence identity. Normalization between Ethanol and Glucose datasets was done by normalization over the mitochondrial dry weight (mtDW). To normalize, the input number of cells in OD units was first normalized to cell dry weight (CDW) using the following relations (experimentally determined as described by Verduyn *et al.* [36] using 30 samples of different OD_{660}):

$CDW_{Ethanol}$ (g L^{-1}) = 0.1257 × OD_{660} - 0.0451 (R^2 of 0.922)

$CDW_{Glucose}$ (g L^{-1}) = 0.1353 × OD_{660} + 0.0451 (R^2 of 0.988)

The CDW was multiplied by the volumes used (Table S3.5), to get the total CDW per condition. Cells have a density of 1 g mL^{-1}, and on ethanol a mitochondrial volume of 0.35 mL mitochondria per mL ethanol-grown

culture and 0.05 mL mitochondria per mL glucose-grown culture [47]. From this follows approximately 0.26 g (± 0.02) mtDW · g CDW^{-1} for ethanol-cultures and 0.027 (± 0.001) g mtDW · g CDW^{-1} for glucose cultures. This conversion factor was used to normalize over mitochondrial dry weight. All counts, either raw, normalized to CPM or normalized to CPM and mtDW are listed in the Supplementary Information (SI_readcounts.xlsx).

Per-base-coverage of the mitochondrial transcriptome was obtained with SAMtools (parameters: samtools depth -aa). Relative per-base-coverage at distinct loci were computed in R (script: coverageplots.rmd) by normalizing coverage at each base position to the total read depth in Counts Per Million (CPM). Ratios between intron and exon levels were determined by calculating the coverage of the last 250 base pairs of each splicing variant or exon, as these base pairs are unique between each variant, and then subsequently normalizing over the exon level.

Read length estimation

For assessment of the read length estimation, a different reference alignment file was used in which the mitochondrial genome was split up into shorter reference sequences that each contained a single mitochondrial open reading frame (ORFs). This allowed for assigning of reads to single ORFs rather than to the full mtDNA, making quantification more straightforward and filtering out erroneous reads with gaps that span multiple ORFs. Read mapping was then performed using the same parameters as described. Using the resulting BAM files, for each ORF the average length of its respective mapped reads was calculated.

Visualization of read mapping

Read mapping was visualized using Integrated Genome Viewer [58] and Tablet [59]. Mapping of long reads was visualized using the full mitochondrial genome as a reference sequence and corresponding BAM files. Splicing was visualized using separate mitochondrial ORFs as reference sequences with the respective BAM files, as this increases resolution of the data when looking into splicing events.

Characterization of intron-less strains

Fluorescence microscopy

Microscopy was performed using a Zeiss Axio Imager Z1 (Carl Zeiss AG, Oberkochen, Germany) equipped with a HAL 100 Halogen illuminator, HBO 100 illuminating system and AxioCam HRm Rev3 detector (60 N-C 1'' 1.0×) (Carl Zeiss AG). The lateral magnification objective 100×/1.3 oil was used with Immersol 518 type F immersion oil (Carls Zeiss AG). Fluoresence of ymTurquoise2 was detected using filter set 47 (Carl Zeiss AG; excitation bandpass filter 436/20 nm, beamsplitter filter 455 nm, emission filter 480/40 nm). Results were analysed using the Fiji package of ImageJ [60]. Mean fluorescence per cell was analyzed by thresholding the phase-contrast channel of each image using the Otsu algorithm to determine the location of the cells in the image. Based on the thresholded phase-contrast image, a binary mask was generated indicating the regions of interest (ROI, i.e. cell locations) of the phase-contrast image. Subsequently, the mask was overlaid on the DAPI channel of the image and the mean fluorescence for each ROI was determined, resulting in a mean fluorescence per cell.

Growth rate analysis

Growth rate analysis was performed in 96-wells microtiter plates at 30 °C and 250 rpm using a Growth Profiler 960 (EnzyScreen BV, Heemstede, The Netherlands). Frozen glycerol stocks were inoculated in 100 mL YPD medium and grown overnight. 0.5 mL of the overnight culture was transferred to 100 mL SMD + Ura/Lys/Ade and grown until the OD_{660} had doubled at least once to ensure exponential growth. From this culture a 96-wells microtiter plate (EnzyScreen, type CR1496dl) containing either SMD + Ura/Lys/Ade or SMEG + Ura/Lys/Ade with final working volumes of 250 µL was inoculated with a starting OD_{660} of 0.1. Growth rate analysis and data analysis was performed as described in Boonekamp *et al.* [41].

Respiration assay

Specific rates of oxygen consumption were measured in 4 mL volume in a stirred chamber at 30 °C. The

oxygen concentration was measured with a Clark-type oxygen electrode (YSI, Yellow Springs, OH). Strains were grown overnight in SMD + Lys/Ade. The overnight culture was diluted twice in 100 mL SMD + Lys/Ade and SMEG + Lys/Ade. Glucose cultures were grown for another three hours to ensure exponential growth, ethanol cultures were grown for another 48 hours to ensure that strains were fully respiratory.
Strains were spun down and washed twice to prevent carry-over of carbon-containing medium and resuspended in SM. For respiration assays on glucose, 25 OD_{660} units of culture were inoculated in a total volume of 4 mL in fully aerated SM + Lys/Ade to a final OD_{660} of 6.25. Air flow in the chamber was stopped and 20 µL of a 50 % (w/v) glucose solution were added to a final concentration of 14 mM. For respiration assays on ethanol 40 OD_{660} units of culture were inoculated in a total volume of 4 mL in fully aerated SM + Lys/Ade to a final OD_{660} of 10. Air flow in the chamber was stopped and 50 µL absolute ethanol solution were added to a final concentration of 210 mM. The percentage of dissolved oxygen was measured until oxygen was depleted and was used to calculate the rate of oxygen consumption expressed as the decrease of oxygen in the reaction vessel over time, divided by the biomass concentration in the vessel.

DNA Sequencing and bioinformatics

Whole genome sequencing of strains 161-U7 and 161-I^0 was performed by Macrogen Europe (Amsterdam, the Netherlands). Genomic DNA of samples 161-U7 and 161-I^0 were sequenced at Macrogen on a Novaseq 6000 sequencer (Illumina, San Diego, CA) to obtain 151 cycle paired-end libraries with an insert-size of 550 bp using TruSeq Nano DNA library preparation, yielding 2 Gigabases in total per sample. Reads of both samples were mapped using BWA (version 0.7.15) [61] against a CEN.PK113-7D genome. Alignments were processed using SAMtools (version 1.3.1) [55]. *De novo* assembly was performed using SPAdes (version 3.9.0) [62] for both samples. Assembled contigs were aligned with nucmer (MUMmer package v3.1) [63] to the CEN.PK113-7D reference, the overlapping contigs aligned to the mitochondrial chromosome were merged into scaffolds to reconstruct the mtDNA for both samples. The contigs containing *COX1* and *COB* as well as the consensus sequences of the mitochondrial genomes of 161-U7 and 161-I^0 were aligned to the CEN.PK113-7D genome using BLAST (NCBI, USA) and visualized in R using genoPlotR [64]. The sequencing data and *de novo* assemblies are deposited at NCBI (https://www.ncbi.nlm.nih.gov/) under BioProject ID PRJNA902953.

3.3 Results

Isolation of mitochondria and preparation of RNA

As compared to nuclear DNA-encoded transcripts, the quantification of all RNA species in mitochondria presents specific technical challenges that require tailor-made experimental and computational procedures. Firstly, the cellular abundance of mitochondrial RNA in yeast is very low and only represents circa 5 % of the total cellular RNA pool under respiratory conditions, and approximately 1 – 3 % under glucose repression, of which 90 – 95 % represents mitochondrial rRNA [25, 65]. Secondly, the absence of polyadenylation of yeast mitochondrial RNA prevents their enrichment by standard, poly(A)-tail based methods. Finally, to fully capture the diversity in mitochondrial RNA splicing variants, processing of the RNA should be kept to a bare minimum, as PCR or cDNA synthesis reactions might introduce biases. To overcome these problems, a mitochondrial RNAseq workflow was developed and tested (Figure 3.2A). Enrichment for mitochondrial RNA was achieved by physically separating mitochondria from yeast cells and extracting RNA from isolated mitochondria. To minimize steps that might bias the mitochondrial RNA pool, either by enriching for specific species or by modifying transcripts length, Nanopore long read direct RNA-sequencing was used, in which a single (polyadenylation) step is required between RNA extraction and sequencing. Synthetic control RNAs were spiked at three different stages to monitor mitochondrial RNA quality during extraction, polyadenylation, and sequencing (Figure 3.2A).
Yeasts cultures were spheroplasted and homogenized, and mitochondria were isolated from the other cellular fractions by differential centrifugation. Oxygen uptake measurements and microscopic analysis confirmed that isolated mitochondria were abundant and intact (Figure S3.1). Western blotting and enzyme assays further showed that mitochondria were significantly enriched, but not fully devoid of cytosolic

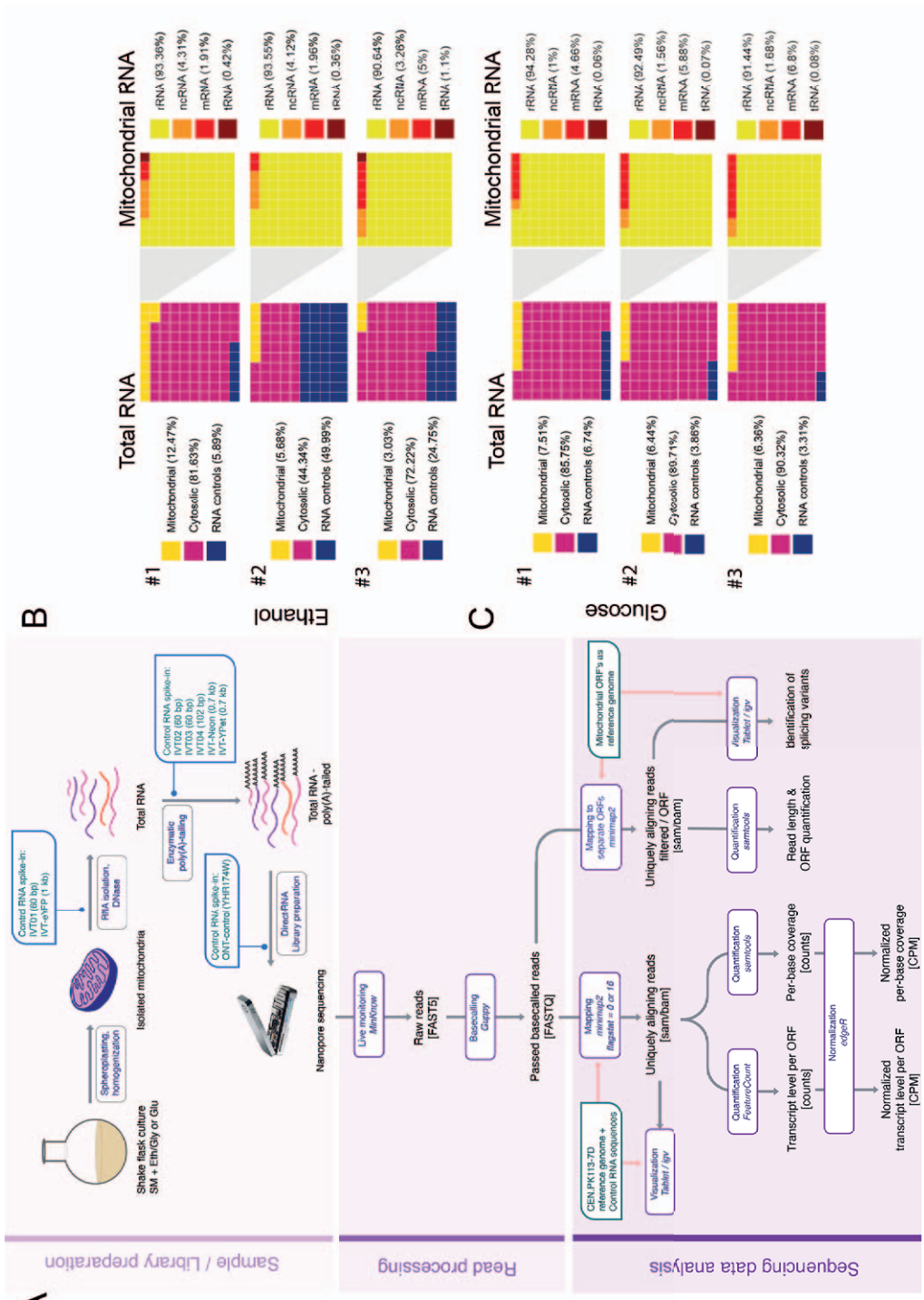
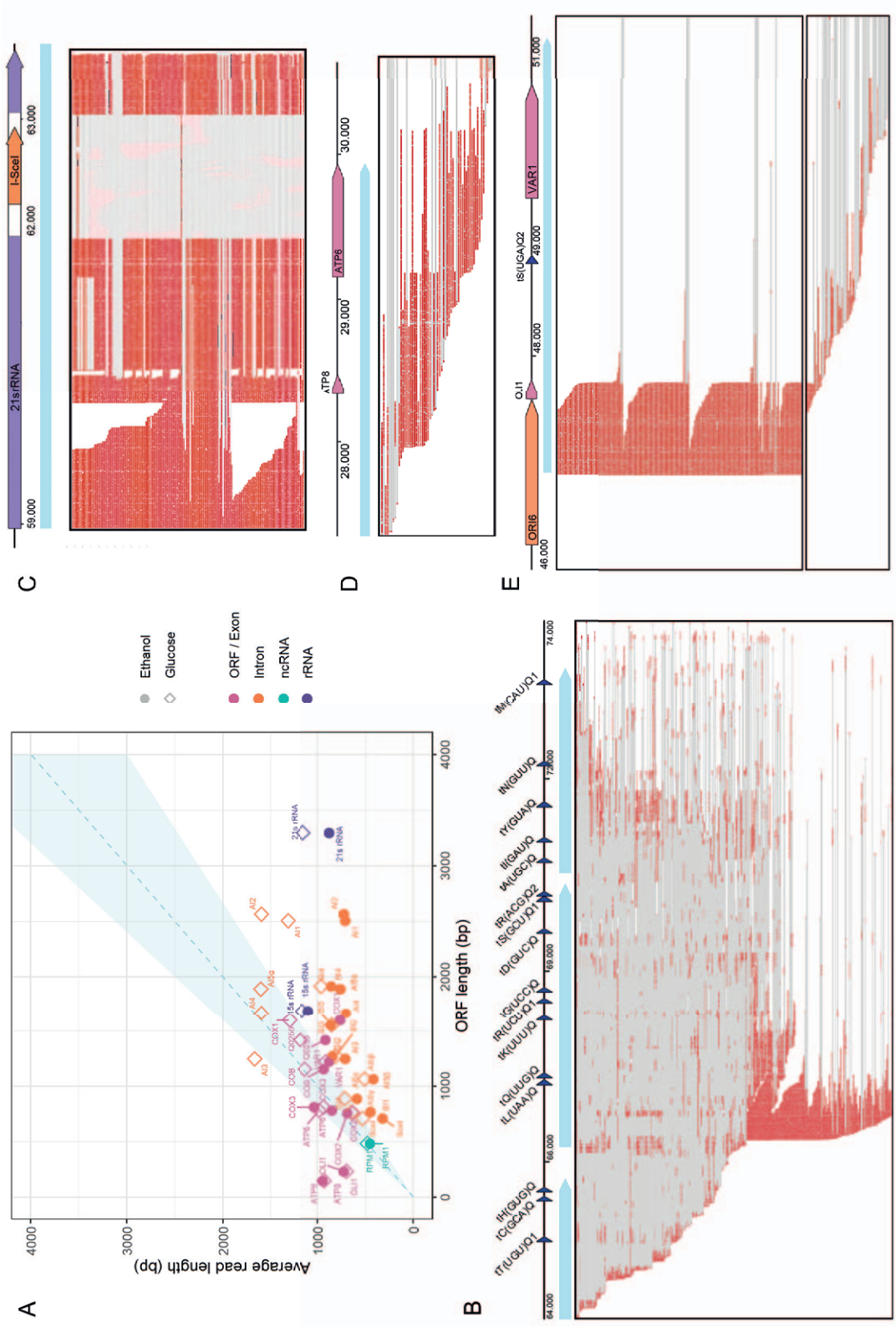


Figure 3.2. Sequencing pipeline and quality control. A) schematic overview of the sample preparation and sequencing pipeline. B, C) Waffle plots showing distribution of read origin and type of the passed reads (Q >= 7) of the three replicates of ethanol-grown cultures (B) and glucose-grown cultures (C). The distribution of the origin of sequencing reads between mitochondrial, cytosolic and control RNA is shown on the left, on the right the distribution of different types of mitochondrial RNA is shown. One square represents 1/100 of the total number of reads.


A
Average read length (bp)
ORF length (bp)
Ethanol
Glucose
ORF / Exon
Intron
ncRNA
rRNA
B
C
21s rRNA
I-SceI
59.000
62.000
63.000
D
ATP8
ATP6
28.000
29.000
30.000
E
ORI6
Q.I1
tS(UGA)Q2
VAR1
46.000
48.000
49.000
51.000

contamination (Figure S3.1). The mitochondrial fraction can be contaminated by carry-over of cytosolic RNA species suspended in the cytosol and by RNA species attached to the outer mitochondrial membrane. These contaminants could be removed by RNAse treatment and by removal of the mitochondria outer membrane (i.e. mitoplasting). These treatments did however not significantly reduce the contamination by most cytosolic RNA and reduced mitochondrial RNA yields (Figure S3.2). RNA was therefore directly extracted from crude isolated mitochondria without additional processing.

To quantify the contamination by non-mitochondrial RNA species and monitor RNA quality during extraction, two synthetic non-polyadenylated control RNAs were spiked to mitochondria prior to RNA extraction. A short (60 bp) *in vitro*-synthetized control simulated a tRNA-length transcript (IVT01), while the 1 kb eYFP transcript simulated messenger RNAs (IVT-eYFP, Figure 3.2A). Control RNAs of various lengths were also spiked further down the workflow. Spiking before poly(A)-tailing assessed the IVT reaction yield and the poly(A)-tail length, and a control supplied by Oxford Nanopore Technologies (ONT) was added prior to sequencing to assess sequencing throughput, and library size and quality (Figure 3.2A). Polyadenylation was required to make the mitochondrial and *in vitro*-synthesized RNA compatible with the Direct RNA sequencing protocol of ONT, as the adapters required for Nanopore sequencing are added through a 10(dT) primer sequence that hybridizes with the poly(A)-tail of mRNA. As mitochondrial RNA is not natively polyadenylated, addition of a poly(A)-tail of at least ten adenine residues was required for successful adapter ligation and subsequent sequencing the RNA.

Following the above-described protocol, RNA-seq was performed using independent triplicate cultures of the *S. cerevisiae* laboratory strain CEN.PK113-7D grown with glucose or a mix of ethanol and glycerol (further referred to as ethanol media) as sole carbon source. RNA sequencing resulted in library sizes ranging between 0.4 to 1.1 Gb of data, with $0.4 \cdot 10^6$ to $1.2 \cdot 10^6$ reads that passed a quality score of above 7 (Table S3.5). The average read N50 score was around 1.2 kb, indicating that most of the reads were of a length corresponding to a mature mRNA. The longest reads were around 8 kb, revealing that some, but not all full-length polycistronic primary transcripts were detected (expected size 1.6 to 16 kb, Figure 3.1). ONT Direct RNA sequencing does sequence poly(A)-tails of the RNA molecules, thereby enabling the quantification of poly(A)-tail length. The enzymatically added poly(A)-tails of mitochondrial and control RNAs were on average 40 - 50 adenine residues long, a length comparable to cytosolic RNA species (Figure S3.3). 95 % of the basecalled reads passed the ONT quality control (Q-value >7) and all passed reads mapped to the reference genome. The mitochondrial RNA species represented around 5 - 12 % of the total number of passed reads, excluding the fraction of control reads , this represented a ca. twofold enrichment of mitochondrial RNA in both conditions, the rest being mostly cytosolic ribosomal RNA (Figure 3.2B,C, Figure S3.4). The libraries were nevertheless large enough to comfortably cover the mitochondrial genome: the obtained datasets had on average $5 \cdot 10^4$ reads that mapped to the mitochondrial genome, equating approximately 600 Mb of sequencing data to cover the 86 kb mitochondrial genome. On ethanol, the breadth of coverage was at an average of 75 % of the heavy strand and 15 % of the light strand of the mitochondrial genome, and all the open reading frames (ORFs) were covered by sequencing reads. Direct RNA sequencing therefore yielded enough reads of quality and of the expected size and was further explored to characterize the mitochondrial transcriptome.

Read distribution and length

A custom bioinformatics pipeline was set up to analyze the basecalled reads. In a first step, reads were aligned to the fully annotated CEN.PK113-7D nuclear and mitochondrial reference genomes (Figure 3.2). All expected RNA types from both mitochondrial and cytosolic origin were identified: mRNA, tRNA, rRNA, and noncoding RNA (ncRNA) which includes the RNA component of RNaseP *RPM1* and RNA sequences required for RNA-primed mitochondrial origins of replication (*ori*). Transcripts were categorized by type and

Figure 3.3. Read length and coverage of genes. A) Average read length for each gene plotted versus the expected ORF length, where the ORF length is defined as the length of a matured RNA from start- to stop codon, after splicing (if applicable). Reads from ethanol grown cultures are shown as closed circles, reads from glucose-grown cultures as open diamonds. Reads are separated per type, mRNA: messenger RNA (pink), ncRNA: non-coding RNA (green) rRNA: ribosomal RNA (purple). Spliced genes are separated between the main exon (pink) and the different spliced-out introns (orange). The dashed line represents a full-length read (gene length = read length), the blue shaded area represents an interval where the read length is 75 % to 125 % of the full gene length. B, C, D, E) Visualization of raw reads of one representative replicate on ethanol, aligned using Tablet with the primary transcripts of the tRNA cluster (B), 21S rRNA (C), *ATP8-6* (D) and *OLI1-VAR1* (E). The colored arrows indicate the genes and their respective locations, the primary polycistronic transcripts are indicated with light blue arrows. Boxes show visualization with Tablet of sequencing reads mapped to the different loci, a red line indicates correct mapping of a sequencing read to the location on the mtDNA, a grey line indicates a gap in the read.

ORF based on their mapped location on the genome, allowing the quantification and analysis of the origin and type of each sequenced transcript (Figure 3.2B). Most mitochondrial reads (90 – 94 %) belonged to rRNA species. All expected rRNA, mRNA and ncRNA were detected, although some in low abundance.

The mitochondrial transcriptome is characterized by a large diversity in length, the shortest being a 71 bp tRNA and the longest a 16.5 kb polycistronic transcript. The length and sequence of transcripts varies depending on processing of polycistronic transcripts and splicing of the respective genes. With the exception of very short RNAs, long-read RNAseq has the capability to capture this mosaic RNA landscape. For most mRNAs, the average read length was within 75 – 125 % of the expected ORF length, meaning that mRNAs were mostly sequenced in a single read (Figure 3.3A, Figure S3.5, Figure S3.5). Slightly shorter read length might reflect partial degradation of the RNA, while longer reads can reflect incomplete RNA processing or the presence of untranslated leader regions on the RNA. Read lengths were plotted against the mRNA length for identification of these longer and shorter reads. A few reads were substantially shorter or longer than the expected ORF length. Notably, the read length for the *ATP8* and *OLI1* genes was respectively six and three times longer than the expected mRNA length. (Figure 3.3A). In line with earlier reports, closer inspection of the read alignment of these loci confirmed the identification of a bicistronic transcript encoding *ATP6* and *ATP8* [66]. However, long *ATP8* transcripts were also due to the presence of 5' untranslated leader (UTL) and 3' untranslated region (UTR) sequences, rather than polycistronic expression with the surrounding genes: 44 % of the reads mapping to *ATP8* were bicistronic, whereas 55 % of the reads were extended due to attached UTL and UTR regions (Figure 3.3D). *OLI1* is part of a 4.8 kb polycistron, however while a few transcripts extending past the *OLI1* dodecamer were detected (Figure 3.3E), no full-length primary transcripts were found. In agreement with previous reports, *OLI1* transcripts elongation was caused by the presence of a 5'UTL [13, 67].
Non-coding RNA sequences were in general shorter than their expected length. The most extreme case was 21S rRNA for which less than 1 % of the reads mapping to 21S rRNA gene displayed the expected gene length. Since the used ORF lengths already exclude intron sequences, this cannot be the result of splicing. Analysis of the raw reads mapped to the 21S rRNA locus revealed that reads were strongly truncated or degraded at approximately 1 – 1.5 kb after the 5' start of the 21S rRNA gene (Figure 3.3C). In addition, 21S rRNA is the longest RNA present in the mitochondria, and may therefore be more sensitive to RNA degradation. Additionally, most of the intron sequences of *COX1* and *COB* (*aI1-aI5γ* and *bI1-4* respectively) were shorter than their expected length. This could potentially be explained by the rapid turnover of spliced introns, which is essential for efficient protein expression and respiration and to prevent intron accumulation [68, 69], but this requires further investigation.

The fraction of tRNAs in the present dataset was particularly low (Figure 3.2B). Processed tRNAs have a short length, ranging from 71 to 90 bp. The data showed that the spiked RNA controls below 102 bp were strongly depleted during the RNA-seq protocol (Figure S3.7). This result was in line with earlier reports that Nanopore sequencing cannot accurately analyze short nucleotide sequences of < 100 bp [70] and suggests that short RNAs such as tRNAs were underrepresented in the present study. All tRNAs are part of a polycistronic transcript, either in a tRNA cluster or grouped with other RNA transcripts. Analysis of reads aligned to the tRNA cluster revealed that single processed tRNAs were indeed rarely detected and that any captured tRNAs were part of a polycistronic transcript (Figure 3.3B).

Based on the nonanucleotide sequences that initiate translation in the mitochondria [12], it is assumed that the mitochondrial genome is expressed in 11 primary polycistronic transcripts that are further processed at internal dodecamer sites and by endolytic cleavage of tRNAs. None of these full length polycistronic transcripts were detected, the longest read was 8 kb and mapped to the *COX1* pre-mRNA. However, knowledge on processing of these polycistronic transcripts is still very limited, and polycistronic transcription has only been demonstrated directly for few primary transcripts [66, 67]. It was only recently found that the dodecamer is recognized by the protein Rmd9 and that deletion of this protein results in detection of unprocessed primary transcripts [71, 72]. Also, processing may be co-transcriptional and therefore difficult to capture *in vivo*. Nevertheless, the read length analysis revealed some discrepancies between read length and gene length that suggest polycistronic expression of the mitochondrial genome.

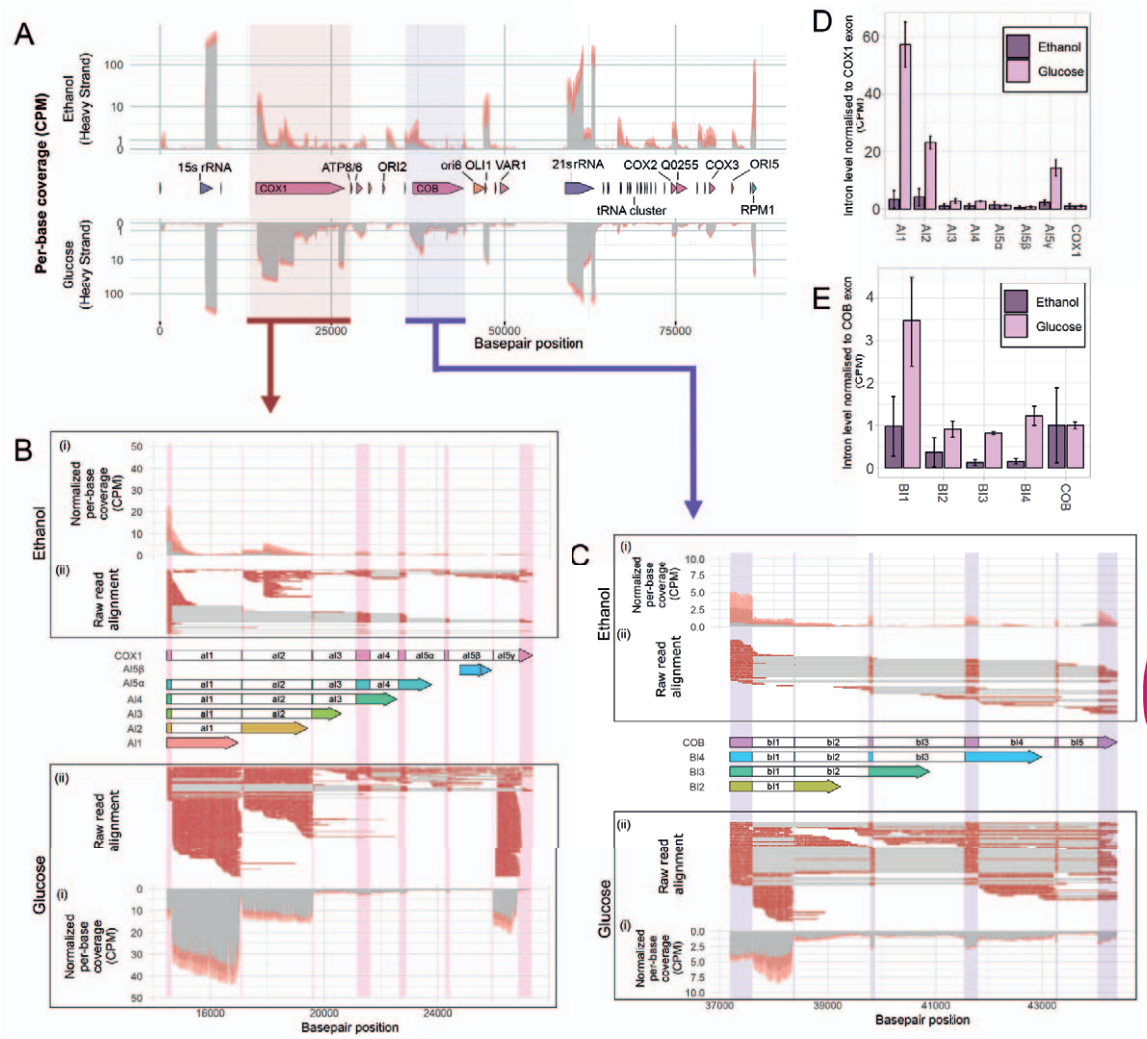


Figure 3.4. Sequencing coverage of the mitochondrial transcriptome of cultures grown on ethanol and glucose. A) Per-base coverage of the mitochondrial transcriptome of cultures grown on ethanol (top) compared to the per-base coverage of the mitochondrial transcriptome of cultures grown on glucose (bottom), normalized per million bases as Counts Per Million (CPM) of the heavy (forward) strand of the mitochondrial DNA. Coverage depth is represented in grey, the standard deviation between the sequencing depth of triplicate experiments is shown in orange. B) Schematic overview, sequencing depth and splicing of the *COX1* open reading frame (ORF). The middle shows the *COX1* ORF, where the exon sequences are shaded in pink, colorless areas are spliced out of the final *COX1* mRNA. The different splicing variants are shown below the *COX1* ORF. Capitalized names are used to indicate protein-encoding introns, non-coding introns are shown in lowercase. (i) shows per-base coverage of the *COX1* ORF, normalized per million bases as Counts-Per-Million (CPM). (ii) shows visualization of sequencing reads mapped to the *COX1* ORF visualized using Tablet, a red line indicates alignment of a sequencing read to the location on the ORF, a grey line indicates a gap in the read. The analysis was done for mitochondria of ethanol-grown cultures (top) and glucose-grown cultures (bottom). The location of the *COX1* exon is shaded in pink for visualization purposes. C) schematic overview, sequencing depth and splicing of the *COB* ORF. D, E) Transcript levels of the *COX1* (D) and *COB* (E) introns on glucose (light pink) and ethanol (purple), normalized to library size in CPM and normalized to transcript levels of the exon mRNA

mtRNA coverage and landscape

The present datasets originated from three biological culture replicates, mitochondria isolation rounds and sequencing runs, leading to different read distributions and library sizes (Figure 3.2B,Table S3.5). Quantitative analysis of expression levels therefore required library normalization. RPKM (Reads Per Kilobase Million) is the standard normalization method for RNA-seq, developed for short-read RNA sequencing in which read length (ca. 150 - 200 bp) is substantially shorter than gene length (i.e. several kb in yeast). RPKM normalizes

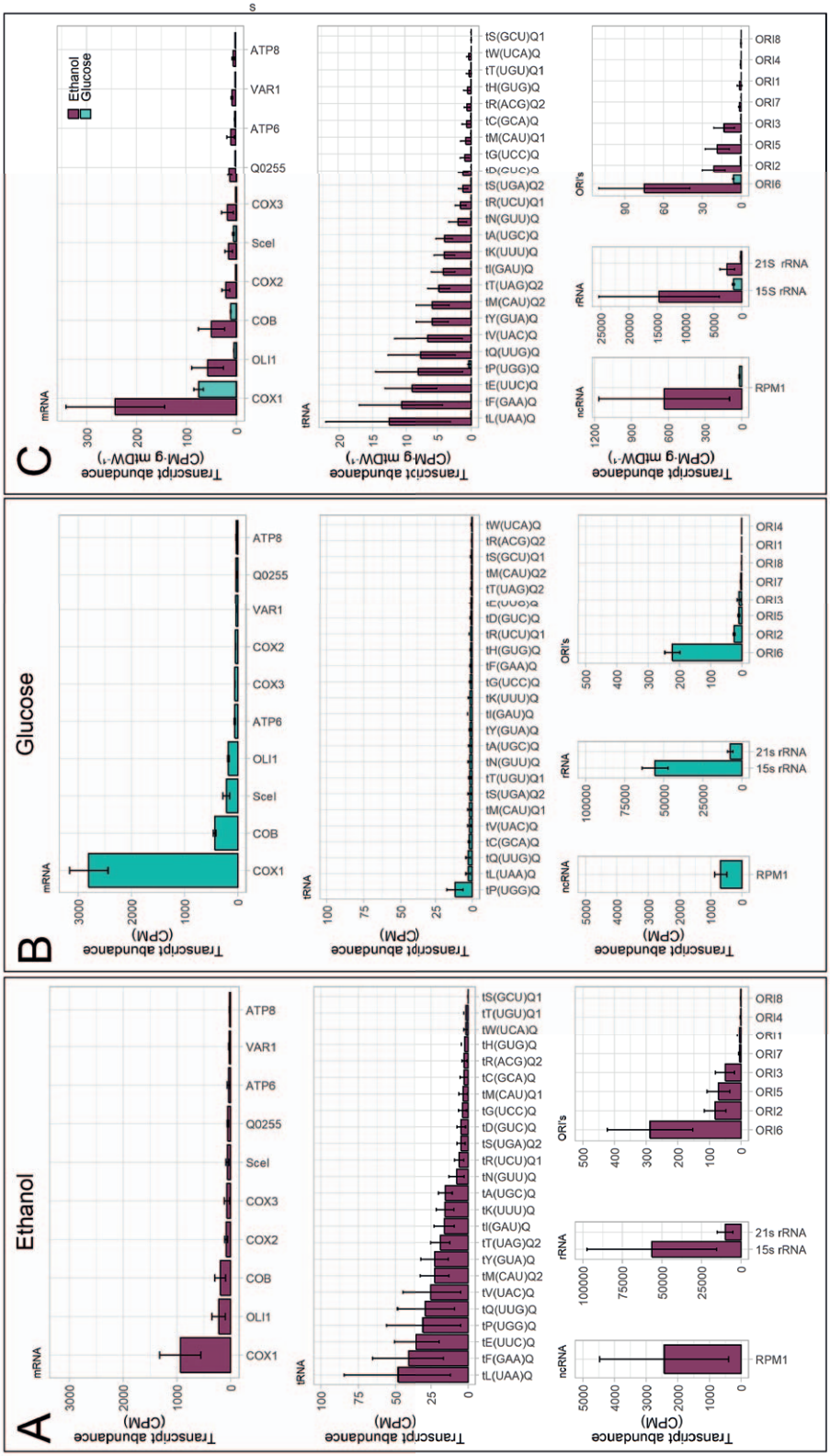


Figure 3.5. Transcript abundance of mitochondrial RNA species. Counts were normalized to library depth in Counts-per-Million (CPM) of mitochondria isolated from ethanol-grown cultures (A) and glucose-grown cultures (B). Error bars indicate the standard error between triplicate experiments. Reads are grouped per type, mRNA: messenger RNA (no separation made between introns and exon reads, these are accumulated in these numbers), ncRNA: non-coding RNA, rRNA: ribosomal RNA, ori: replication origin. C) Comparison of transcript abundance of mitochondrial RNA species of mitochondria isolated from ethanol-grown cultures (dark purple) and glucose-grown cultures (light blue), normalized to library depth in Counts-per-Million (CPM) and normalized to grams mitochondrial dry weight of the cultures (g mtDW).

to library size, but also to gene length, as to compensate for this direct correlation between read number and gene length. However, normalization by gene length is not applicable to long-read RNA sequencing as read length often approaches or even exceeds the gene length. The datasets were therefore only normalized over library depth in Counts Per Million (CPM).

The sum of polycistronic primary transcripts of the mtDNA encompassed 54.1 kb of the heavy strand, which means that 64 % of the heavy strand is theoretically expressed, and 6.3 kb (7 %) of the light strand. In the obtained RNAseq dataset, the breadth of coverage of the mitochondrial DNA heavy strand was higher than expected, with 74.7 % and 85.7 % for ethanol- and glucose-grown cells respectively (Figure 3.4A,B, Figure S3.8, Figure S3.9), revealing a very good representation of the mitochondrial transcriptome. When looking at the raw read alignment of coding and non-coding regions, the respective 10 – 20 % higher expression does not have a clear origin and seems to mostly result from a combination of occasional read-through past the dodecamer or mis-alignment of read ends (Figure S3.10), which was especially prevalent in the AT-rich 3' end of 15S rRNA (Figure S3.11).
For the light DNA strand, only containing one active origin of replication and tRNA tT(UAG)Q2, the observed coverage was 14.8 % for ethanol-grown and 7 % for glucose-grown cultures (Figure 3.4A, Figure S3.8, Figure S3.9). Most expression clearly originated from the expected primary transcripts. On ethanol, the breadth of coverage of the light strand was slightly higher than expected, but the depth of this coverage was very low. This might result from transcriptional read-through at the *ori3* locus or from the presence of mirror RNAs, which have been described to exist in similar low quantities in both human and yeast mitochondria and are possibly a by-product of RNA processing [13, 73].

On ethanol, all of the coding regions were sufficiently covered by sequencing reads (Figure 3.4, Figure S3.10), including regions encoding tRNAs, albeit in a low amount and often as polycistronic reads (Figure 3.3B). The gaps in coverage represented 14 – 25 % of the mitochondrial genome, but all in non-coding regions. The mitochondrial genome was therefore exhaustively sequenced, despite the low representation of mitochondrial RNAs in the total RNA-seq data (*ca.* 7 %, Figure 3.2B). Irrespective of the carbon source used, the per-base coverage showed a large variation in expression between the mitochondrial genes (up to 3300-fold difference, Figure 3.4A). The described method yielded a complete coverage of the mitochondrial genome and could therefore be used to quantify changes in RNA expression and processing of the mitochondrial genome between different conditions.

Comparison of transcript levels between ethanol and glucose-grown cultures

While comparing the relative expression of the various mitochondrial RNA species within a given growth condition is straightforward and only requires normalization over library depth in Counts Per Million (CPM), comparing RNA abundance between growth conditions is more challenging. *S. cerevisiae* is a typical Crabtree positive yeast, in which excess glucose triggers a respiro-fermentative metabolism. Glucose exerts a strong repression on respiration, leading to low respiration rates, small and sparse mitochondria, and lower mitochondrial RNA levels [25, 74, 75]. Conversely, ethanol is fully respired, leading to abundant and very active mitochondria. The present method first extracted mitochondria from different culture volumes on glucose and ethanol, while RNA sequencing is performed on the same input amount of RNA. Comparing mtRNA levels during growth on glucose and ethanol therefore requires an additional correction for the abundance of mitochondrial mass per cell, between the two carbon sources. Since direct RNAseq results in smaller library sizes (~ 0.5-1 M aligned reads) well below the required input to obtain a high-power differential gene expression using short-read sequencing (~ 10 M reads), gene expression was measured as transcript-level counts [22, 76, 77]. Additionally, a recent study reported a correlation between mitochondrial mass and cell dry weight of glucose-grown and ethanol-grown cultures [47]. This correlation was used to normalize the counts per mitochondrial dry weight, thereby enabling the direct comparison of mtRNA abundance between glucose- and ethanol-grown cultures (Figure 3.5, Table S3.5).

In both conditions, based on non-normalized data, the most abundant RNA species were rRNAs, the 15S rRNA being 5.5 times more abundant than the 21S rRNA (Figure 3.5A, B). Transcripts of all mitochondrial mRNAs were recovered in both conditions, but their relative expression was affected by the type of carbon source. For instance, transcripts mapping to the cytochrome *c* oxidase subunits 1, 2 and 3 (*COX1*, *COX2* and *COX3*, respectively), the cytochrome *b* (*COB*) and the F0-ATPase subunit c (*OLI1*) were most abundant in ethanol-grown cultures (Figure 3.5A). Conversely, on glucose, transcripts mapping to the *COX1* ORF were most abundant, followed by *COB* and the meganuclease-enconding transcript *I-SceI*, encoded in the omega-intron of 21S rRNA (Figure 3.5B). In line with the reduced involvement of mitochondria in respiration with excess glucose, the overall normalized abundance of mtRNA was lower in glucose than in ethanol-grown cultures. Accordingly, with the notable exception of *COX1* and *COB*, the abundance of transcripts encoding respiratory chain subunits was 5 to 15 times higher in ethanol- than in glucose-grown cultures (Figure 3.5C, Figure S3.12). The high expression of *COX1* relative to *COX2* in glucose conditions was unexpected. *COX1*, *2* and *3* together encode the catalytic core of cytochrome *c* oxidase and exist in a 1:1:1 stoichiometry in the inner mitochondrial membrane [78-80]. The presented finding also contradicts a previous study on the stoichiometry of mitochondrial protein- and RNA levels under glucose repression [25, 47].
Additionally, tRNAs were more abundant in ethanol than in glucose-grown cultures, which is in line with the increase in protein synthesis in ethanol-grown cultures and previous studies on tRNA levels under glucose repression [81]. Out of the 24 tRNAs, 22 could be quantified in all ethanol-grown cultures, while only tP(UGG)Q was reliably detected in glucose-grown cultures (Figure 3.5). In line with these observations, the third most abundant RNA species in ethanol-grown cultures was *RPM1*. *RPM1* encodes the RNA component of RNase P, an enzyme highly active during respiratory growth that is essential for tRNA processing [82, 83]. All tRNAs are part of a polycistronic transcript, either in a tRNA cluster or grouped with other RNAs, and the three most abundant RNA species tL(UAA)Q, tF(GAA)Q and tE(UUC)Q are located at the start of a polycistronic primary transcript (Figure 3.1).

The dataset also provides insight on the expression of the origins of replication. It is still unclear whether mitochondrial DNA replication relies on RNA priming, rolling-circle replication, or a combination of the two [84, 85]. Nevertheless, transcripts for three (*ori2*, *ori3*, *ori5*) out of the eight mitochondrial origins of replication were detected under both growth conditions, suggesting that these origins were active under the tested conditions (Figure 3.4A, Figure 3.5). This is accordance with the presence of uninterrupted promoter sequences for these *ori*s [13, 86-88]. The abundance of these replication primers was three- to five-fold higher with ethanol as carbon source (Figure 3.5). Many reads also mapped to *ori6*, which was most likely an artefact caused by the presence of the *OLI1*/*VAR1* primary transcript within this origin. This hypothesis was supported by the lack of reads mapping to the 5'-end of *ori6* (Figure 3.3E).
Remarkably, the read coverage within genes was also carbon-source dependent, particularly for the *COX1* and *COB*, and 21S rRNA loci (Figure 3.4A). For instance, the coverage within the 21S rRNA on ethanol as compared to glucose shows a sudden decrease in between the 5' end and the omega-intron, at approximately 1 kb after the 5' end, which corresponds to the previously observed gap in read alignment (Figure 3.3C). This gap suggests that reads might be truncated or degraded from this point onward. The fact that the dip in coverage originated from the same position suggests that shearing of RNA during processing is not a likely cause, as this RNA damage would occur randomly. A similar dip in coverage was observed for the *RPM1* gene, for which the decrease in coverage and read truncation exactly matched the RNA processing pattern and proposed heptakaidecamer cleavage site (Figure S3.13, [13]). Taking this into consideration, the coverage pattern observed for 21S rRNA might suggest the presence of an RNA processing site.

Long read sequencing can therefore be used for the quantitative comparison of transcript levels between growth conditions. However, the polycistronic nature of the long RNA reads requires careful interpretation of the data taking into consideration the mapped location and coverage within a gene of the reads.

Changes in splicing patterns of *COX1* and *COB* between growth on ethanol and glucose

Comparison of the read coverage on glucose and ethanol-grown cultures showed pronounced differences for *COX1* and *COB* (Figure 3.4A-C), with a 5 to 50-fold increase in coverage of sub-sections of these ORFs when glucose was the carbon source. A uniform coverage over the full ORF is not expected, since the *COX1* and *COB* ORFs do not only encode subunits of cytochrome *c* oxidase and cytochrome *b*, respectively but also contain several introns and intron-encoded genes (Figure 3.4B,C). The *COX1*-ORF (or pre-mRNA) is 12.9 kb long, of which 1.6 kb consist of exon sequences that together encode Cox1p. The *COB* pre-mRNA is 7.1 kb long of which 1.2 kb encode Cobp. The *COX1* and *COB* introns belong to either group I or group II homing introns that are distinguished by their splicing mechanism and are considered phenotypically nonessential for *S. cerevisiae* [20, 89, 90]. The *COX1* exon sequences are interrupted by four group I and three group II intron sequences (*COX1-aI1*, *aI2*, *aI3*, *aI4*, *aI5α,β,γ)*, and *COB* contains one group I and four group II intron sequences (*COB-bI1*, *bI2*, *bI3*, *bI4*, *bI5)*, that are spliced out of the pre-mRNA in a protein-facilitated manner [20, 35, 91]. As a result of alternative splicing (part of) these introns can also encode proteins (Figure 3.4B,C). *COB-BI2*, *-BI3* and *–BI4* encode maturases, *COX1-AI1* and *–AI2* encode reverse transcriptases, and *COX1-AI3*, *-AI4* and *-AI5α* encode endonucleases, all of which facilitate splicing and intron mobility of their respective intron sequence [8, 20, 91, 92]. Finally, *COX1-AI5β* encodes a putative protein of unknown function, group II introns *aI5γ* and *bI1* are non-coding.

In ethanol-grown cultures, the pre-mRNAs of both *COX1* and *COB* were clearly spliced and processed to yield an RNA containing only exon sequences (Figure 3.4B,C). For *COX1*, spliced intron sequences were also detected, but displayed various degrees of degradation and were rarely found in full-length, in agreement with the shorter than expected read length of intron sequences (Figure 3.3A). Overall, no significant difference was observed between intron and exon abundance of *COX1* in ethanol-grown cultures (Figure 3.4B,D). For *COB*, intron sequences were rare, especially introns *bI3* and *bI4* were strongly depleted and present at only 10 % of the exon level. Some reads showed patterns of alternative splicing, but since intron levels are relatively low and intron turnover is co-translational in mitochondria [93], it is difficult to attribute the intron levels to either 'regular' (*aI1-5*, *bI1-5*) or alternative splicing (*AI1-5*, *BI2-4*) of the intron.
In glucose-grown cultures, the splicing pattern of the mature *COX1* and *COB* exon was more difficult to infer from the raw read alignment, and unspliced pre-mRNAs were clearly more abundant (Figure 3.4B,C). Also in contrast with ethanol-grown cultures, in glucose-grown cultures the coverage of several sequences mapping to intronic regions was substantially higher than the corresponding exon abundance for both *COX1* and *COB* (Figure 3.4B,C). *COX1-aI1*, *-aI2* and *-aI5γ* were respectively 57-, 23- and 14-times more abundant than the mature *COX1* exon-RNA (Figure 3.4D). Similarly, the *COB-bI1* intron was 3.5-times more abundant than the *COB* mature exon-RNA (Figure 3.4E). Di Bartolomeo *et al.* [47] reported that the abundance of the Cox1 protein was 15 to 50 times higher than its intron-encoded proteins, a result in contrast with the relative exon-intron abundance measured in the present study (Figure S3.14). Additionally, analysis of the raw read alignment to the *COX1* and *COB* ORFs revealed that the increased intron levels were not due to alternative splicing, as the intron sequences were rarely attached to a 5' exon sequence (Figure 3.4B,C). Lastly, considering that the *aI5γ* and *bI1* introns do not encode any protein [20], the high intron abundance most likely results from the accumulation of the spliced intron sequences rather than the elevated expression of intron-encoded genes.

Strikingly, all of the introns with increased levels on glucose (*COX1-aI1,2,5γ*, *COB-bI1*) belong to the group II introns, while the levels of group I intron sequences (*COX1-aI3,4,5α,5β* and *COB-bI2,3,4*) remained similar to the level of exon RNA (Figure 3.4D,E). Group II introns are circular RNAs (known as lariats) and are a result from the covalent linking between the 5' end and an A/U-residue 8 to 9 base pairs from the intron 3' end. The formation of this branch point is followed by exon ligation and cleavage at the 3' splice site and the resulting release of a circular intron lariat [20, 94, 95]. Nanopore technology cannot sequence circular RNA molecules, which means that (part of) the lariats were most likely linearized prior to sequencing, either *in vivo* or during sample processing. Introns *aI1* and *aI2* are 2.3 kb in length, while *aI5γ* and *bI1* are both around 0.8 kb. Over 75 % of reads for the *COX1-aI1*, *-aI2*, *aI5γ* and *bI1* intron had the expected length up to the branch point, which indicates little to no degradation (Figure 3.4B,C). Additionally, the median weighted read

length (N50) of the dataset was 1.4 kb and most of the RNA was sequenced in full (Figure 3.3A), meaning RNA degradation during library preparation was likely limited. Lastly, mechanic shearing of the circular RNA during library preparation would result in a random location of the break for each RNA molecule. When comparing the alignment of the raw reads to each intron, no random breakage of reads was observed, as reads either consistently covered the full intron from 5' to 3' end or were partially degraded from the 5' end (Figure 3.4B,C).
This data suggested the accumulated group II introns of mitochondria of glucose-grown cells were not sheared and were sequenced in full, up to the branching point. Approximately half of the reads for these introns included the branch point sequence, with only very few (5 to 10 reads per sample) surpassing the branch point (Figure S3.15, Figure S3.16, Figure S3.17, Figure S3.18). Additionally, the abundance of the group II intron sequences is likely underrepresented in both glucose- and ethanol-grown cultures, as circular lariat molecules were most probably present in the isolated RNA, but were not available for sequencing due to the nature of Nanopore technology.

Exploring the effect of the presence of group II introns on yeast physiology

Group II introns specifically accumulated during respiro-fermentative growth in glucose medium. While this accumulation might be a side effect of some yet unknown condition-dependent regulation of intron stability, the possibility of a physiological role for mitochondrial lariats cannot be excluded. In mammalian cells, there are indications that noncoding RNA can play a role in gene expression, epigenetic regulation, and genome stability [96-98]. The hypothesis that intron sequences may play a similar role in yeast can be tested by comparing the physiology of yeast strains with and without group II introns. Despite the poor genetic accessibility of the mitochondrial genome, a few 'intron-less' strains have been previously constructed [35, 99]. The physiological response of an intron-less strain from the 161 strain lineage (also known as ID41-6/161) to growth in the presence of glucose or ethanol was therefore explored [33, 34]. The intron-less strain 161-I^0 does not contain introns in *COX1* and *COB* and also lacks the ω-intron (encoding I-SceI). *COX1* and *COB* introns were removed through biolistic transformation ([35, 99], personal communication). The control congenic strain 161-U7 only lacks the ω-intron.

As the sequence of these strains' genome was unavailable, *S. cerevisiae* 161-U7 and 161-I^0 genomes were sequenced using Illumina short-read technology. A *de novo* assembly was performed based on the sequencing reads (Table S3.6), and resulting contigs containing *COX1* and *COB* could be aligned to the CEN.PK113-7D mtDNA to determine the absence of intron sequences (Figure 3.6A). Additionally, a consensus sequence of the mitochondrial genomes of 161-U7 and 161-I^0 was generated by mapping the contigs from the *de novo* assembly of the two strains to the CEN.PK113-7D mitochondrial genome and extracting the resulting consensus sequence (Figure S3.19). In addition to the mitochondrial genome, the complete sequence of the nuclear genome of these strains was *de novo* assembled and is available at NCBI (see M&M section).
Comparing the mitochondrial consensus sequences of the 161-U7 and 161-I^0 confirmed the absence of the ω-intron in both strains (Figure S3.19) and the absence of all other introns in strain 161-I^0 (Figure 3.6). The mtDNA of 161-I^0 covered 77 % of the 161-U7 genome and the similarity between the two mitochondrial consensus genomes was 99.2 % (Table S3.7). The *de novo* assembly resulted in fragmented mitochondrial genomes (Table S3.6) and must be complemented with long-read sequencing to confirm the correct structure of the mtDNA, therefore the analysis of the *COX1* and *COB* loci was performed using the correctly assembled *de novo* contigs that contained these loci rather than the generated consensus sequence (Figure 3.6). The similarity between the *COX1* and *COB* alleles was 99.8 and 99.9 % respectively (Table S3.7), yielding 12 SNPs in *COX1* and one in *COB.* Additionally, in the *COB* locus of intron-less strain 161-I^0, the 14 bp exon sequence between introns *bI1* and *bI2* appeared to be removed from the mtDNA (Figure 3.6A).

A previous study reported substantial physiological differences between the two strains during growth on glucose, but growth on ethanol was not explored [100]. In contrast with this earlier study, 161-U7 and its intron-less variant grew at similar rates on glucose medium. In cultures with ethanol as carbon source the

intron-less strain grew slightly but significantly slower (14 %, p-value 0.05, student t-test homoscedastic) than its parent (Figure 3.6B). Such a change in growth rate may be explained by mutations that occurred in construction of strain 161-I^0 or the absence of the second exon of *COB*. The respiration rate was similar for the two strains during growth on ethanol and was increased by ca. 30 % during growth on glucose for the intron-less strain (Figure 3.6C), which might indicate a partial alleviation of glucose repression. The two strains were equipped with a mitochondrial fluorescent protein to monitor mitochondrial mass. When applied to the CEN.PK113-7D control strain, this method showed a four-fold decrease in mitochondrial mass in glucose-grown as compared to ethanol-grown cultures (Figure 3.6D,E). This result was in good agreement with earlier reports for the same strain [47]. The same method applied to 161-U7 and 161-I^0 showed a substantially smaller mitochondrial mass variation between glucose- and ethanol-grown cells, where mitochondrial mass on glucose represented approx. 60 % of the mitochondrial mass on ethanol but revealed the absence of difference in mitochondrial mass between the two strains. The absence of introns did therefore not lead to clear physiological differences when comparing growth on glucose and ethanol media.

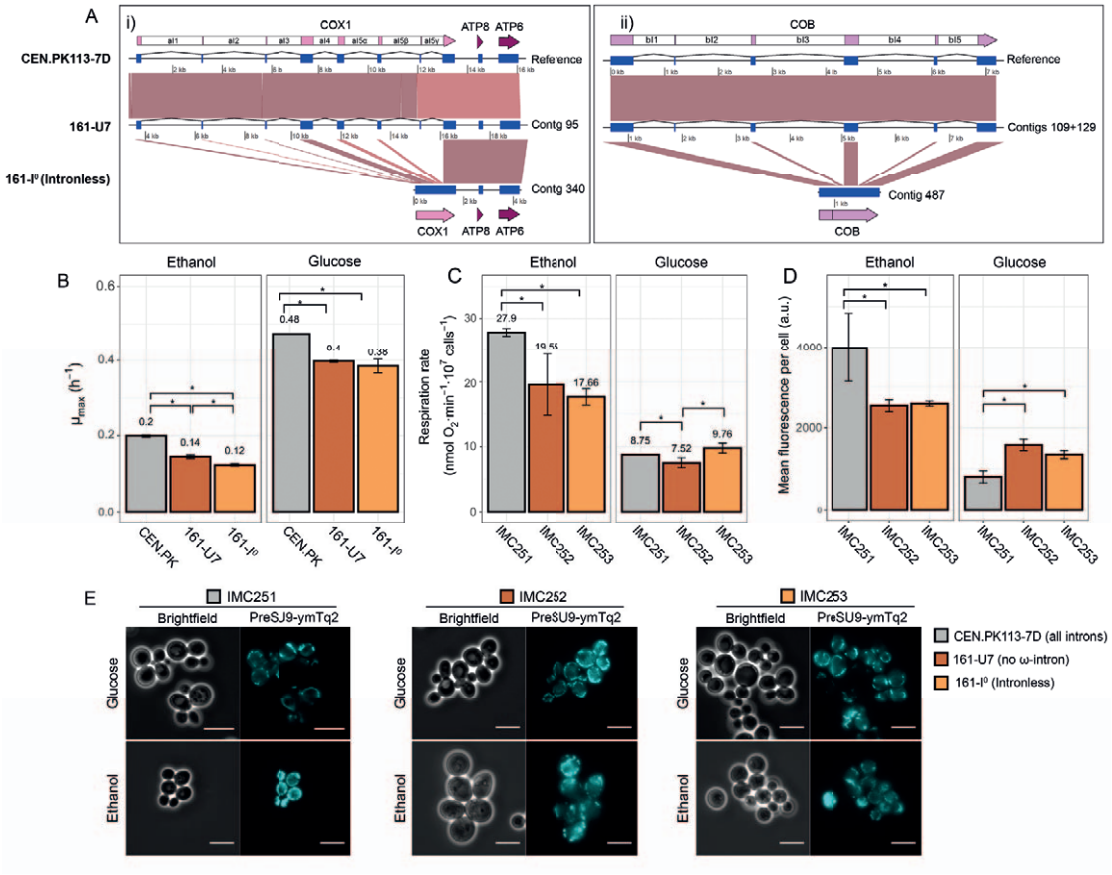


Figure 3.6. Comparison of the genomes and physiology of the CEN.PK113-7D, 161-U7 and 161-I^0 strains. A) Alignment of the i) *COX1* and ii) *COB* loci on the mitochondrial genome of CEN.PK113-7D, and contigs of 161-U7 and 161-I^0. *COB* of 161-U7 was divided in 2 contigs which were combined at a 76-bp overlap. The mitochondrial genome is represented as a black line and the respective coordinates are indicated below each genome/contig. The locations of exon mRNA, rRNA and ncRNA are indicated as blue squares, tRNA and *ori* are not shown. Spliced exons are connected by a line, annotations of the sequences are indicated above (CEN.PK113-7D, 161-U7) and below (161-I^0) the alignment. Grey boxes or lines indicate a sequence identity of > 95 % (BLAST) between the two alignments. B) Maximum specific growth rate (h^{-1}) in microtiter plates of CEN.PK113-7D, 161-U7 and 161-I^0 on ethanol and glucose medium. C) Respiration rate, D) Mean fluorescence intensity per cell measured in the mTurquoise2 channel, determined by microscopy and E) Microscopy pictures of strains IMC251 (CEN.PK113-7D), IMC252 (161-U7) and IMC253 (161-I^0) grown on SM-ethanol or SM-glucose with a blue fluorescent mTurquoise2 (mTq2) protein targeted to the mitochondria with a preSU9 mitochondrial targeting signal. Mean fluorescence was calculated per cell (n = 250), the cell area was based on the brightfield image. Scale bars represent 10 µm. An asterisk (*) represents a significant difference with p < 0.05 (paired two-tailed student T-test).

3.4 Discussion

In the present study, RNA enrichment by mitochondria isolation combined with Nanopore long-read sequencing enabled the capture of the full mitochondrial transcriptome in the model eukaryote *S. cerevisiae*. All expected RNA species were found and the mitochondrial RNA coverage was sufficient to quantify most of the transcriptome. Small RNA molecules of below 100 bp, specifically tRNAs, were an exception, thus studies focusing on these RNA species should rather opt for, or complement with, short read sequencing, or use methods to specifically enrich these transcripts [101]. The coverage of the mtRNA was sufficient, but can be further improved through removing contaminating RNA species by further purification of mitochondria (e.g., using Nycodenz gradient purification [102]). Additionally, treatment with nucleases could reduce carry-over of cytosolic RNA, which represents up to 97 % of the sequenced RNA in the present study. Depletion of rRNA [103] is another approach to enrich samples for the relevant RNA species. Finally, the excessive number of reads originating from the RNA controls can be substantially reduced by decreasing the amount of spiked-in control RNA. Despite the polycistronic nature of the mitochondrial RNA, no full-length primary transcripts were detected. This result might reflect the real *in vivo* situation, in which the short half-life of polycistrons results from co-transcriptional splicing [93]. However, although mitochondrial RNA is stabilized at its dodecamer sequences, it cannot be excluded that during sample processing of the intact mitochondria, the RNA processing and turnover enzymes stay active [23, 71], particularly during spheroplasting which is performed at 30 ˚C for 30 – 45 minutes. Other studies were able to detect primary transcripts, albeit in extremely low amounts [66, 67], so increasing the amount of mitochondrial RNA input by further purification of mitochondrial RNA may also aid in detection of these polycistronic transcripts. Nevertheless, investigating the possibility of additional quenching or stabilization of the mitochondria prior to isolation may improve detection of unprocessed RNA [104, 105].

While the present approach cannot give a perfect snapshot of *in vivo* RNA abundance, by applying the exact same protocol to all samples quantified, mtRNA abundance between different conditions was demostrated with good reproducibility between biological replicates (Figure 3.5). In a previous study, the transcriptome of *S. cerevisiae* grown in ethanol medium was monitored using short read sequencing [13]. The relative abundance of RNA species from short-read sequencing is overall in good agreement with the present study. Small variations can likely be attributed to differences in RNA preparation methods, yeast strain genetic background, sequencing methods and normalization of the data (RKPM by Turk *et al.* and CPM in the present study). Notable exceptions were the expression levels for 15S rRNA and *RPM1*, which were the first and third most abundant species in the presented study, with a 50- and two-fold higher expression compared to *COX1*, whereas Turk *et al.* reported the abundance of 15S rRNA to be comparable to *COX1* and *RPM1* approximately 10-fold lower compared to *COX1*. The data of Turk *et al.* [13] are comparable to an earlier quantification of rRNA by blotting performed by Mueller and Getz [25]. This may be explained by the secondary structure of these RNAs, which can interfere with binding of short sequencing reads or RNA probes, whereas Nanopore direct RNA sequencing includes relaxation of secondary structures [77, 106]. Nevertheless, these large differences between the present study and earlier reports on rRNA level could be an artefact induced by Nanopore sequencing and in future studies, it would be elegant to include a short-read method as a comparison to identify potential artefacts. The levels of mitochondrial rRNA could not be determined by the used ScreenTape assay for RNA quality control, as there was too much contamination by cytosolic rRNA (Figure S3.20).

As opposed to short-read sequencing, the present approach enabled the analysis and visualization of splicing variants with single base pair resolution, and the quantification of the abundance of RNA species at different processing stages. As transcripts were sequenced in a single read, RNA processing events could be identified by combining information on transcript coverage and on the presence of gaps in read mapping to the reference genome. This was best illustrated by comparing the known RNA cleavage sites of *RPM1* with the coverage and read alignment of that gene (Figure S3.13). Similarly, the 21S rRNA showed a strong decrease in coverage and gap in read alignment in a single locus during growth on ethanol media. The consistent occurrence of the same phenomenon for all three independent replicates for ethanol-grown

cultures, but absence for glucose-grown cultures, strongly suggested the existence of a condition-dependent new processing site for these RNA species. As the 21S rRNA sequence harbors no heptakaidecamer or internal dodecamer splicing sites, the exact target site for this novel RNA processing event remains to be defined.

The most remarkable difference between glucose- and ethanol-grown cultures was the substantially higher abundance (4 to 60-fold) of group II introns *COX1-al1*, *-al2*, *-al5γ* and *COB-bI1* in glucose-grown cultures, compared to the exon abundance. Splicing of these lariat introns requires their circularization via internal covalent linking. However, Nanopore technology cannot sequence circular RNA molecules, and the 8-9 bp branch of the lariat is too short to be poly(A) tailed [94], which indicates that the fraction of introns that was sequenced exists in linear configuration. The existence of linear intron RNA in the mitochondria has been hypothesized but never demonstrated *in vivo* [81]. Linearization can result from different mechanisms, either mechanical, such as by shearing during sample processing or through *in vivo* processes, including debranching by specific enzymes, degradation by endonucleases, hydrolysis or intron lariat reopening reactions causing linear intron intermediates [107-109]. The reproducibility of group II introns abundance and length between biological replicates suggests that the difference in intron abundance observed between glucose- and ethanol-grown cultures was not an artefact caused by mechanical RNA shearing during sample processing, as approximately half of the reads covered the lariat branching point and were sequenced in full length (Figure S3.15 - Figure S3.18), The only known intron-debranching enzyme in *S. cerevisiae* (Dbr1p [110]) is not localized to mitochondria (prediction using DeepLoc2.0, Table S3.8 [111]) and Dbr1p has not been detected in any mitochondrial proteome [47, 112]. Therefore, it can be reasonably assumed that lariats are either debranched *in vivo* by a yet unidentified debranching enzyme, or that (a subset of) the group II introns is spliced as linear RNA molecules when grown on glucose. This may also mean that the presented intron levels are an underrepresentation of the actual intron level, as the introns with lariat structure could not be sequenced using Nanopore sequencing. Using *in vitro* debranching with purified Dbr1 could provide more insight on this.

Regardless of the configuration of the group II introns, a clear condition-dependent change in group II introns abundance was observed when cultures were grown on glucose. The most likely mechanisms leading to this change in intron-abundance are an increase in either intron stabilization by proteins on glucose or *in vivo* degradation of introns on ethanol. Potential target proteins may be identified by comparing the differential expression of RNA binding proteins observed between glucose and ethanol-grown cultures [47, 113, 114]. From this comparison, *YBR238C*, mitochondrial paralog of Rmd9p that stabilizes mitochondrial RNA at their dodecamer sequence [71, 72], was 1.8 times more abundant in glucose than ethanol-grown cultures (Figure S3.14). MtRNA binding protein levels between ethanol and glucose were also compared relative to Cox1p expression, as this gives insight in the different levels of protein required between ethanol and glucose to mature a similar level of *COX1* mRNA. When normalizing to Cox1p levels, Mss116p, an RNA-helicase and chaperone involved in splicing group II introns [35], showed over two-fold increase in expression between growth on ethanol and glucose (Figure S3.14), as well as Dis3p, which was demonstrated to play a role in mitochondrial intron decay [13]. Combined with the difference in intron RNA abundance between glucose- and ethanol-grown cultures, this data suggests condition-dependent group II intron processing or stabilization by RNA-processing enzymes. The underlying molecular mechanism remains to be explored.

While accumulation of introns in glucose-grown cultures might be a mere side effect of condition-dependent intron turn-over rates, an interesting fundamental question is whether these group II introns have an impact on mitochondrial functions. In yeast, it has been demonstrated that efficient splicing and regulation of group I introns plays a regulatory role [69, 100, 115], while in mammalian cells, non-coding mtRNA may play a role in cell signalling and mitochondrial expression [97, 98]. Such functions have not yet been described for group II introns. The group II introns are not conserved within the *Saccharomyces* genus [116], and as intron-free strains did not show a different phenotype in ethanol medium, it was assumed that group II introns are phenotypically neutral [89, 90]. Intriguingly, one study reported increased respiration rates and mitochondrial volume during growth on glucose medium using a fully intron-less *S. cerevisiae* strain, a phenotype which could also be achieved by overexpressing the Mss116p RNA-helicase and –chaperone involved in group II

introns splicing [35, 100]. As there is no transcriptome data available for the 161 lineage control during growth on glucose and ethanol, there is currently no evidence of any condition-dependent variation in group II introns levels in this specific lineage. Conversely, intron-less strains from the CEN.PK *S. cerevisiae* lineage, used for transcriptome analysis in the present study, are not available and are challenging to construct. More research will therefore be required by either constructing intron-less CEN.PK strains or by quantifying mtRNA levels in strains of the 161 lineage, to conclude on a potential physiological effect of group II introns.

Little is known about the context-dependency of mitochondrial gene expression and splicing. Using the developed method, the impact of environment on transcriptional responses in the tractable model organism *S. cerevisiae* can be uncovered at a base pair-level resolution, both in terms of RNA quantity and processing under different conditions. Considering the important role of mitochondria in aging in mammals, monitoring transcriptomic responses of yeast mitochondria to oxidative stress or in chronologically aging cultures might reveal new layers in gene expression regulation. The presented methodology therefore opens the door to a deeper understanding of mitochondrial processes and their regulation.

Acknowledgements

The authors thank Prof. Alan Lambowitz and dr. Georg Mohr for providing strains 161-U7 and 161-I^0 and Prof. Alan Lambowitz for feedback on the manuscript.

3.5 Supplementary data

A

	IMC173	
	Succinate	ADP
Oxygen uptake (uM)	67,01	99,06
Time (min)	6,00	5,00
Oxygen uptake rate (uM/min)	11,17	19,81
Coupling ratio (ADP/Succinate)		**1,77**

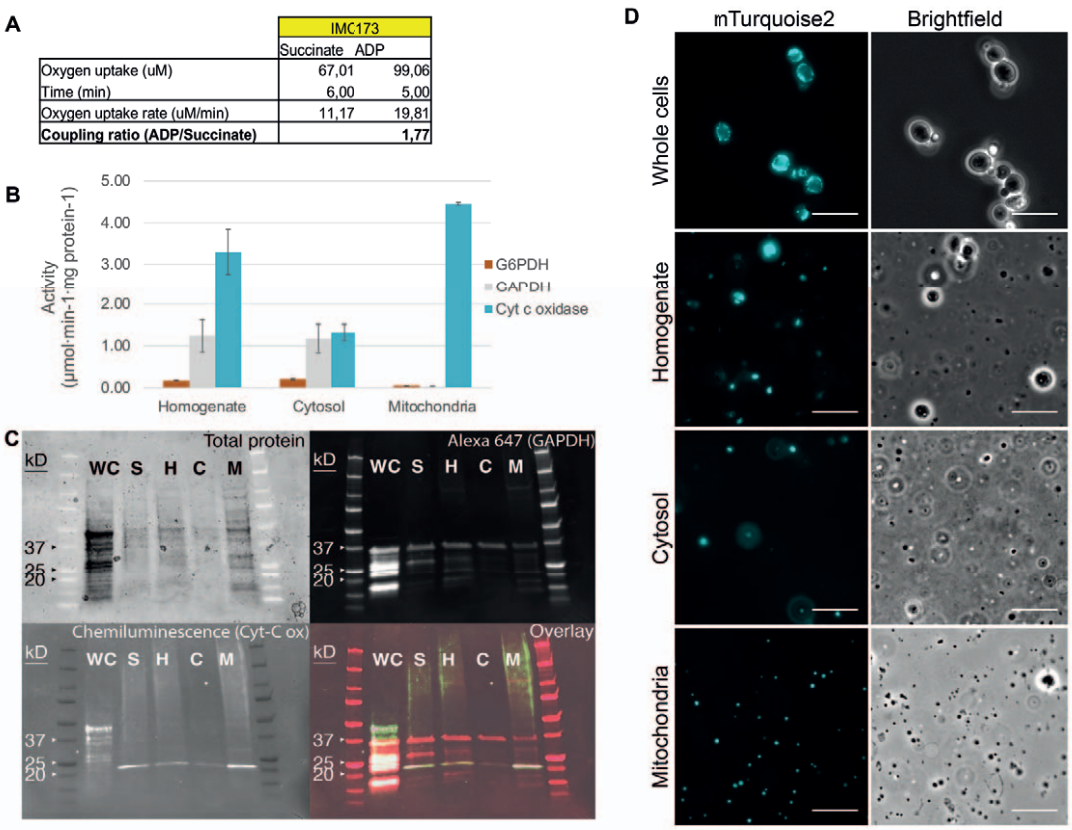


Figure S3.1. Quality assessment of the mitochondria isolation protocol using strain IMC173 (preSU9-mTq2, blue fluorescent mitochondria). A) Oxygen consumption rate of isolated mitochondria supplied first with succinate then with ADP. Details on the experimental setup are described in the supplementary methods. The ratio of the oxygen uptake rates before and after the ADP addition informs on the activity of the respiratory chain and therefore on the integrity of the mitochondrial membranes. The coupling (P/O) ratio when using succinate as substrate is expected to be close to 1.5 [117]. B) Enzyme activity of the cytosolic glucose-6-phosphate dehydrogenase (G6PDH) and glyceraldehyde-3-phosphate dehydrogenase (GAPDH) and mitochondrial cytochrome *c* oxidase (Cyt c, mitochondrial). Activity was measured in different fractions obtained throughout the mitochondria isolation protocol, homogenate: lysed spheroplasts, cytosol: cytosolic fraction after differential centrifugation, mitochondria: mitochondrial fraction after differential centrifugation. C) Western Blots of the mitochondrial fraction. GAPDH (cytosolic, 35,7 kDa) was detected with conjugated anti-GAPDH antibodies containing Alexa 647 and cytochrome *c* oxidase subunit 3 (mitochondrial, 30 kDa) was detected with an anti-Cox3 antibody and a secondary antibody with an HRP group and chemiluminescent substrate. Protein content was imaged using BioRad Stain-Free gels activated under UV. Different fractions of the mitochondrial isolation protocol were analyzed, WC: whole cells, S: supernatant after spheroplasting, H: homogenate (lysed spheroplasts), C: cytosolic fraction, M: mitochondrial fraction. D) Microscopy images of whole cells, homogenate, cytosol and mitochondria. Left: blue fluorescence (ex: 436 nm, em: 480 nm), right: widefield image. Scale bars represent 10 µm

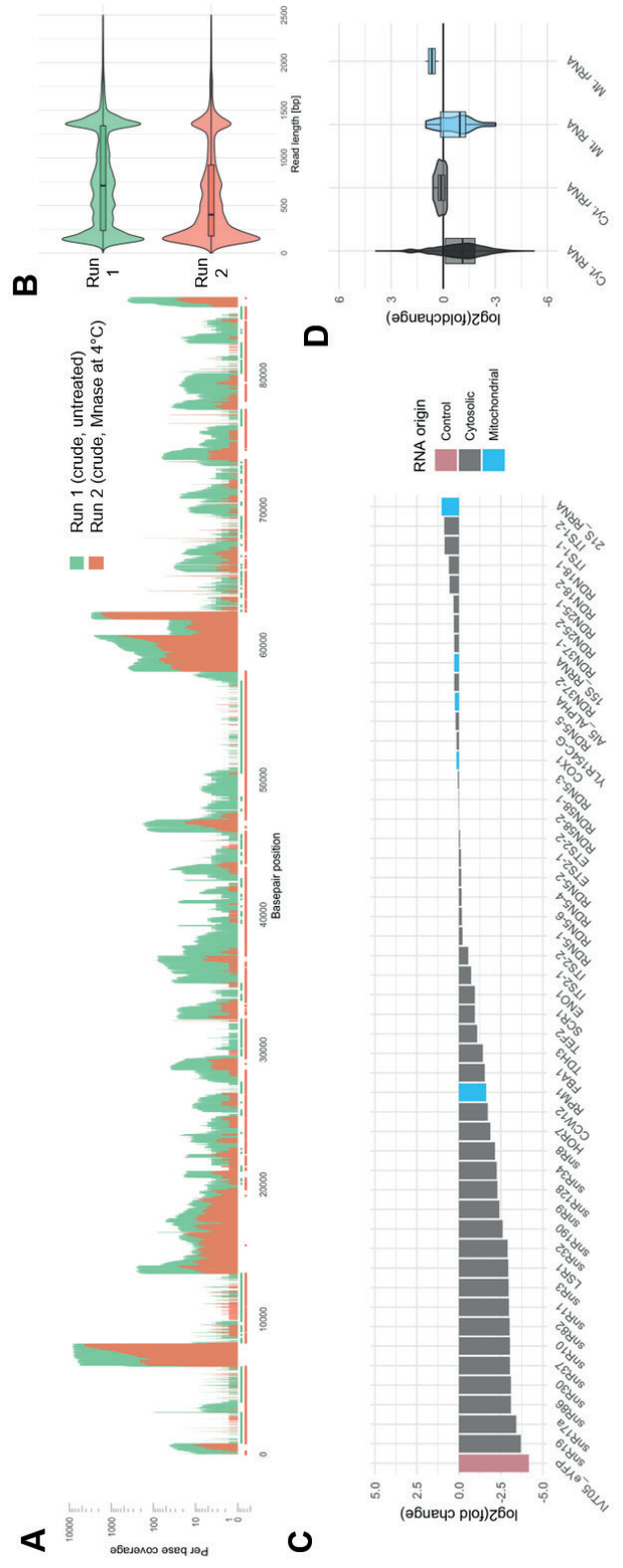


Figure S3.2. Comparison of sequencing data of crude mitochondrial fraction (run 1) and MNase-treated mitochondria (run 2). A) Per-base-coverage of the full mitochondrial genome, with per-base coverage of crude mitochondria (run 1) in green and per-base coverage of MNase-treated mitochondria (run 2) in orange on $\log_{10}$ scale. B) Distribution of obtained read lengths for run 1 and run 2 shown as violin plots, the mean is shown as a box plot where the outer ends of the box represent the 25th and 75th percentile of the data. C) Top 50 most abundant transcripts of run 2 and their fold change of CPM between run 1 and run2, sorted by $\log_2$ fold change in increasing manner. D) Log_2 fold change of CPM of run 1 and 2, grouped by different type (cytosolic RNA except rRNA, cytosolic rRNA, mitochondrial RNA except rRNA and mitochondrial RNA).

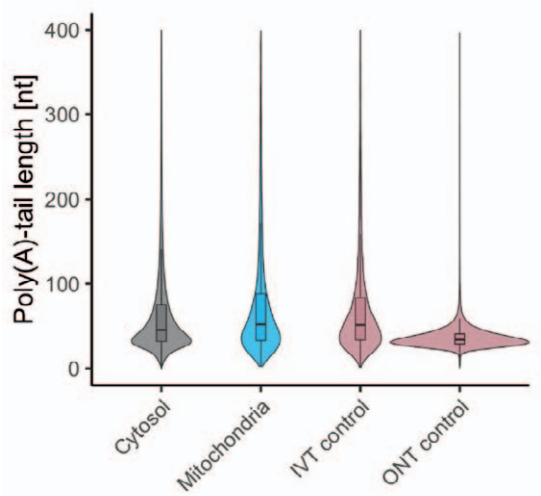


Figure S3.3. Estimated Poly(A)-tail lengths after enzymatic polyadenylation. The length of the poly(A)-tail was estimated by Nanopolish-PolyA on the obtained RNAseq data for RNA of mitochondrial (no native poly(A) tail) and cytosolic origin (natively poly(A)-tailed) and of *in vitro*-synthesized control RNA (no native poly(A)-tail). The length of the poly(A)-tail of the ONT control RNA (not enzymatically polyadenylated, contains poly(A)-tail) was determined as a control. Distribution of poly(A)-tail lengths is shown as violin plots, the mean is shown as a box plot where the outer ends of the box represent the 25th and 75th percentile of the data.

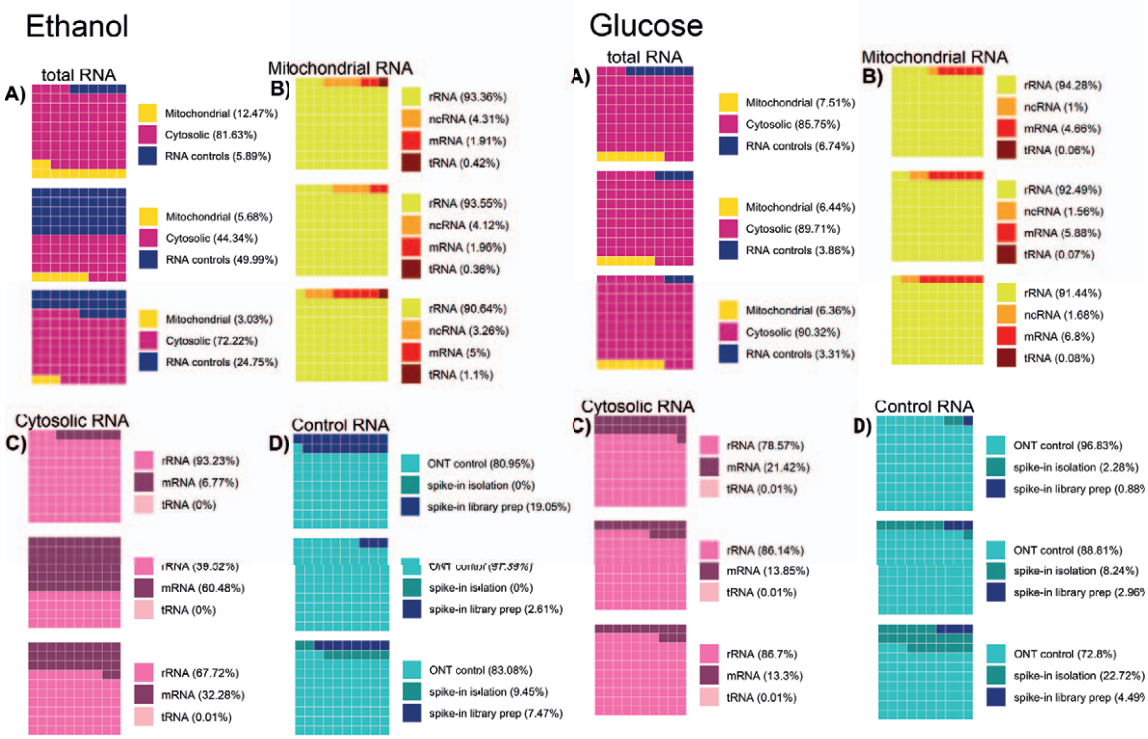


Figure S3.4. Waffle plots of read origin and type for the different sequencing runs of triplicate experiments on glucose and ethanol. A) Distribution of the spatial origin of sequencing reads between mitochondrial, cytosolic and control RNA. Breakdown of the distribution of different types of mitochondrial RNA (B), cytosolic RNA (C), and spiked-in control RNA (D). One square represents 1/100 of the total number of reads. Left panel: ethanol-grown cultures, right panel: glucose-grown cultures.

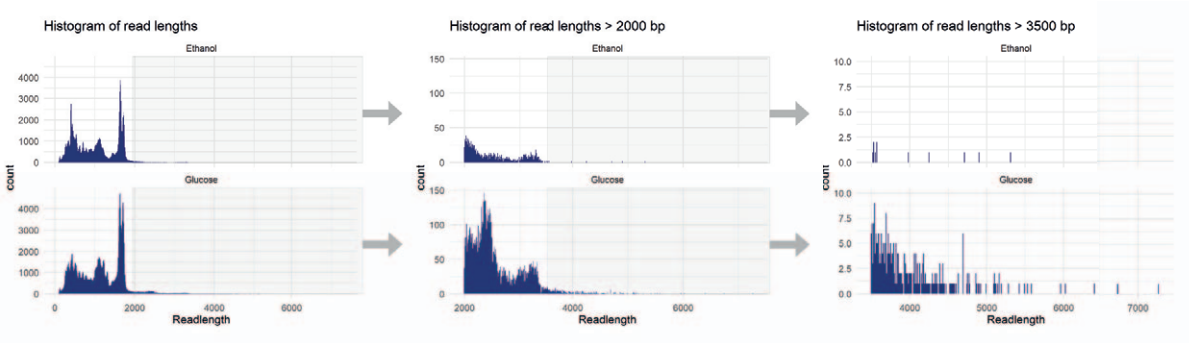


Figure S3.5. Distribution of read counts according to read length in the sequencing dataset. The data show the cumulated read lengths obtained in three replicate experiments on either ethanol (top graphs) or glucose (bottom graphs) as carbon source. A) Distribution of all read lengths. B) and C): Zoom in for the distribution for read lengths larger than 2000 bp (B) and 3500 bp (C) that have lower abundance.

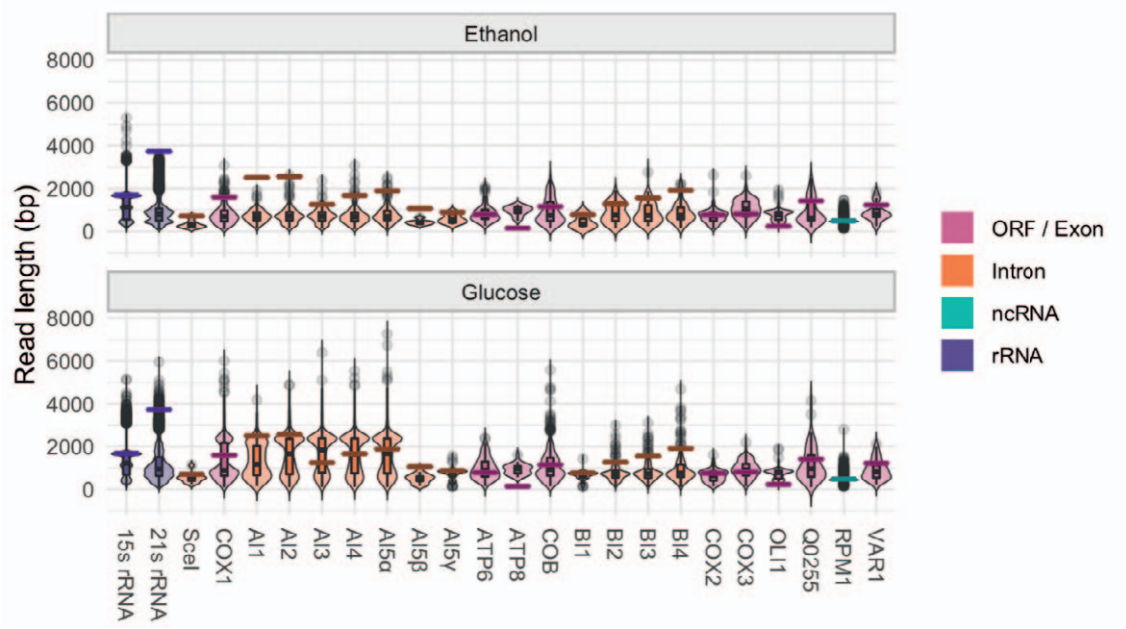


Figure S3.6. Read length distribution per gene. The distribution of the average read length of the triplicate experiments is shown as violins, the average mean read length is shown as a box plot where the outer ends of the box represent the 25th and 75th percentile of the data. Outliers are shown as circles. The gene length is indicated with a horizontal line. Reads are colored per type, mRNA (pink), ncRNA (green) rRNA (purple). Spliced genes are separated between the main exon (pink) and the different spliced-out introns (orange).

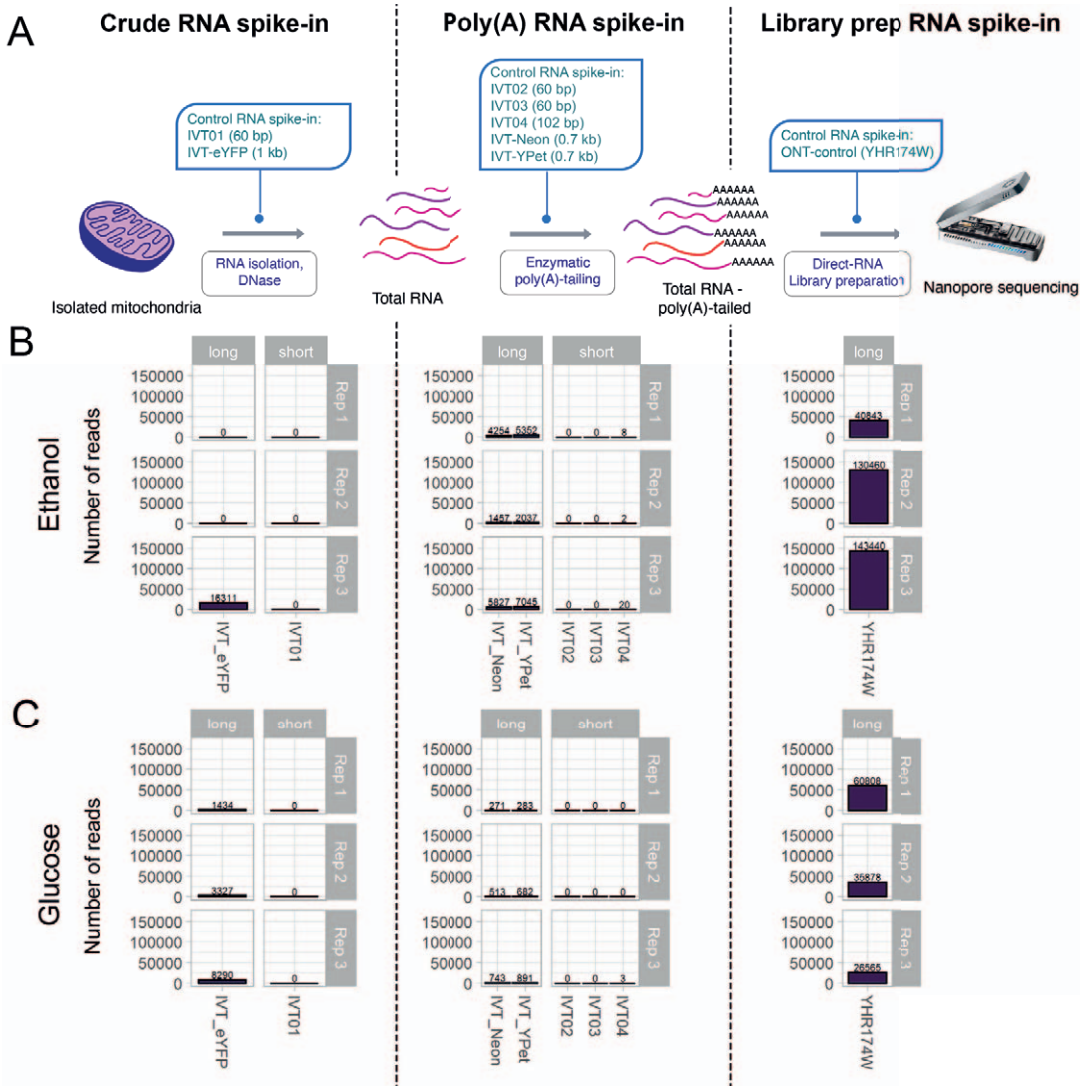


Figure S3.7. Overview of control RNA abundance over all transcriptome experiments. Number of control reads sequenced for all sequenced replicates. A) Schematic overview of the mtRNA sequencing protocol indicating in which steps the control RNAs are spiked. All reads were spiked in equimolar amounts (45 fmol), except for the ONT control read YHR174W of which 61 fmol was spiked according to the ONT direct sequencing protocol. B) and C) Number of sequenced reads for the controls spiked as indicated in A). Data are shown for each independent culture replicate (Rep 1 to Rep 3) and for each spiking step..The absolute number of passed reads in each replicate is plotted without normalization, the numbers above the bars show the number of passed reads per control sequence. 'Long' denotes a read > 150 bp, 'short' reads < 150 bp. Rep = Replicate. B) Cultures grown on ethanol, C) Cultures grown on glucose.

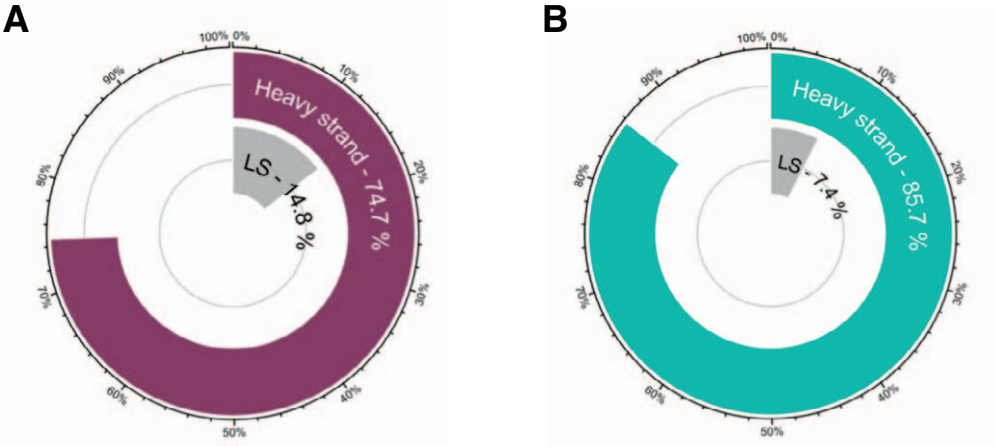


Figure S3.8. Breadth of coverage of the heavy and light strand (LS) of the mitochondrial genome. A) Ethanol-grown cultures. B) Glucose-grown cultures. The coverage is defined as the number of bases that were significantly expressed in triplicate experiments, divided by the size of the mitochondrial genome.

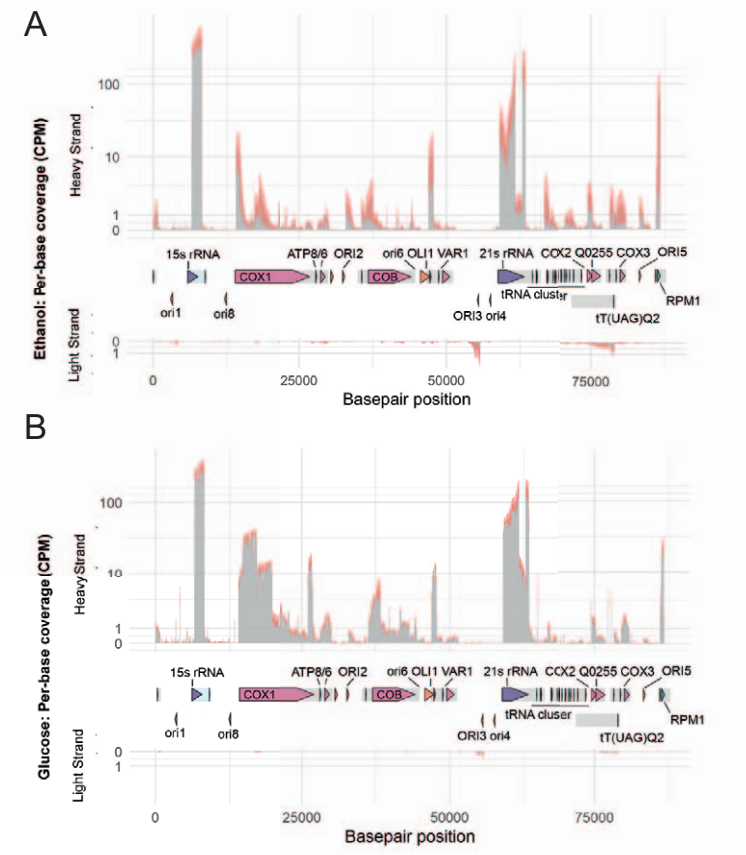


Figure S3.9. Per-base coverage of the mitochondrial transcriptome. A) Ethanol-grown cultures B) Glucose-grown cultures. bottom). The data are normalized per million bases as Counts Per Million (CPM) and both the heavy (forward, top graph) and light (reverse, bottom graph) strands of the mitochondrial DNA are represented. Coverage depth is represented in grey, while the standard deviation between the sequencing depth of triplicate experiments is shown as a red area. The mitochondrial genome is shown between the graphs. Grey boxes indicate primary transcripts.

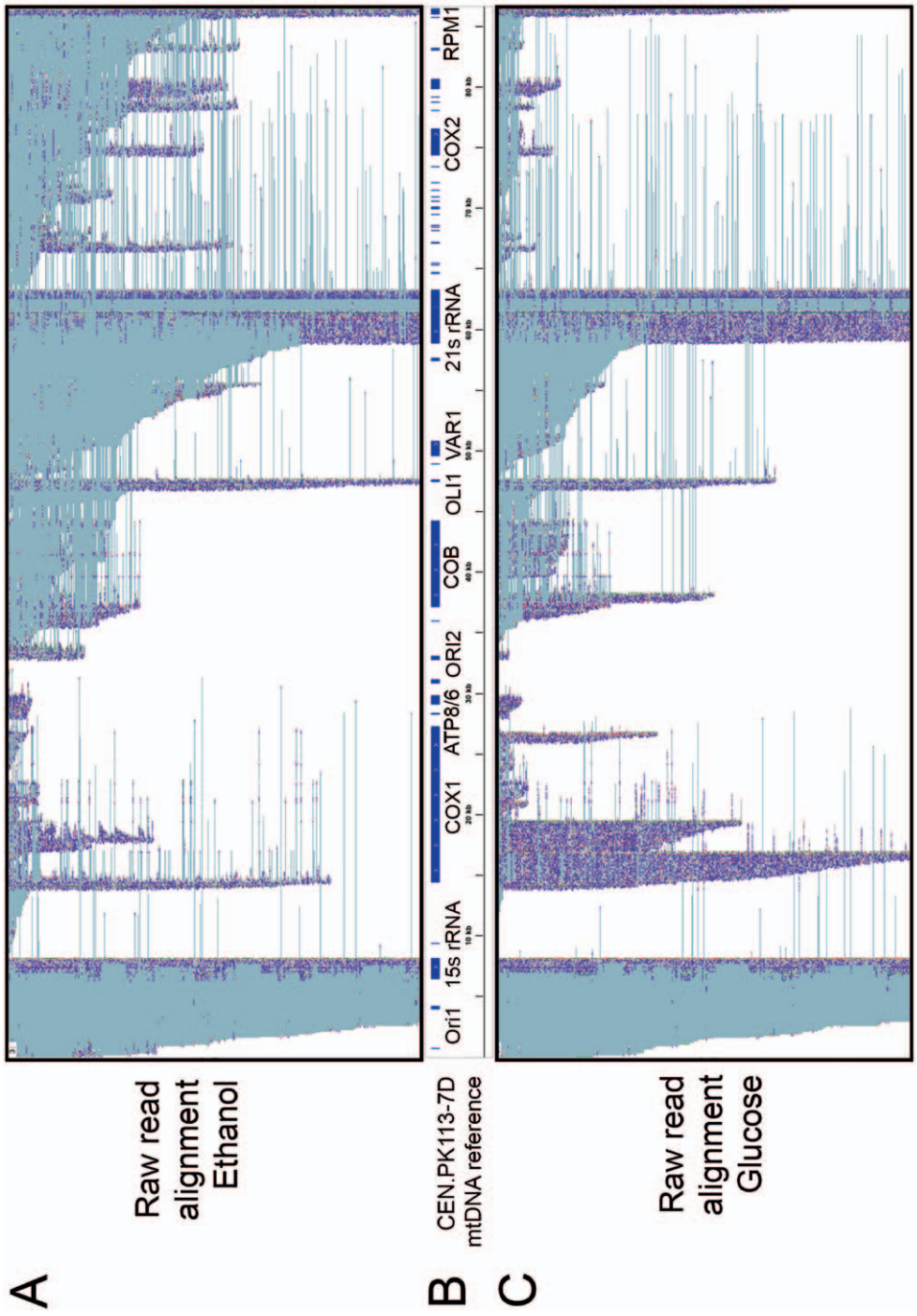


Figure S3.10. Raw read alignment in igv of mitochondrial RNAseq data across the entire mitochondrial genome. A single representative replicate is shown for growth on ethanol (A) and growth on glucose (C). Each line is one read, colors indicate the different bases of the read (A = green, T = red, C = dark blue, G = orange). Grey-blue lines indicate gaps in the reads. B) Annotation of the CEN.PK113-7D mitochondrial reference genome corresponding to the read alignments shown in A) and C).

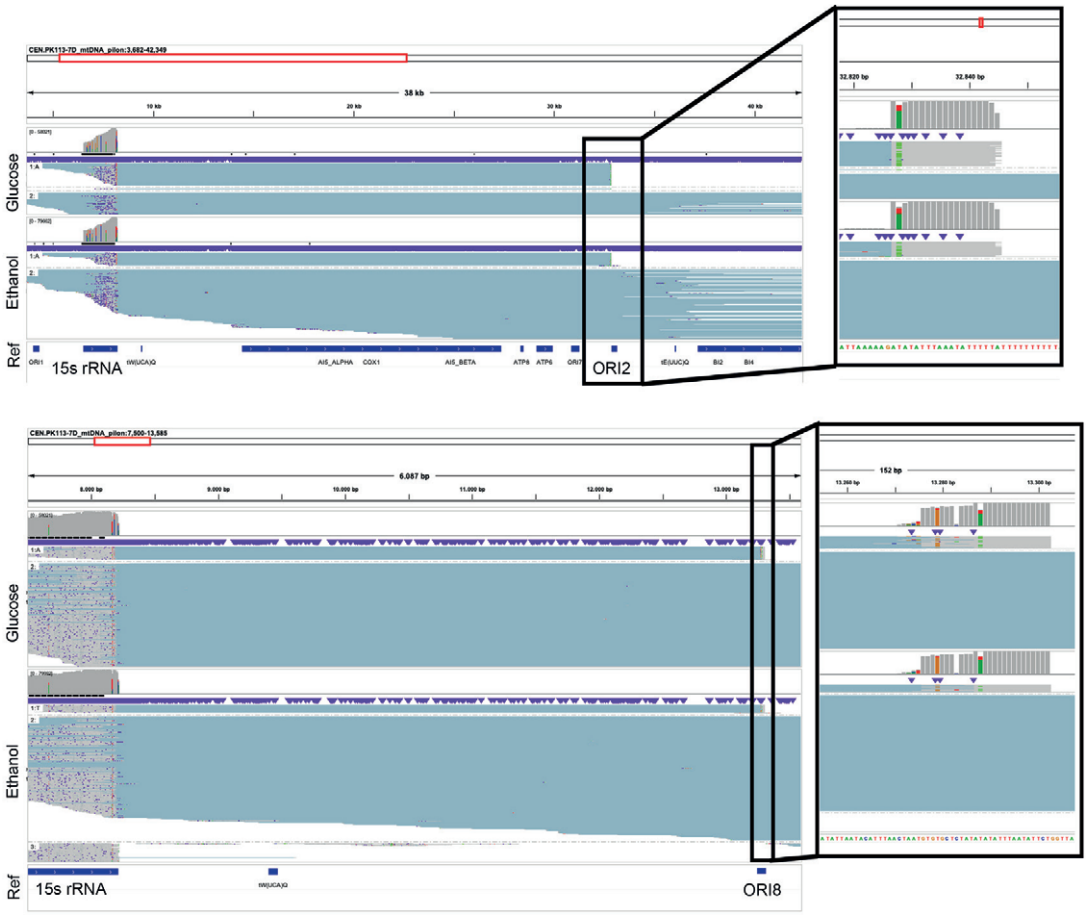


Figure S3.11. Raw read alignment in igv of mitochondrial RNAseq data in the 15s rRNA to *ORI2* (A) and *ORI8* (B) loci. In each panel a single representative replicate is shown for ethanol (bottom graph) and glucose-grown culture (top graph). Grey, dark blue, yellow, green and red bases indicate read mapping, a light blue line indicates a gap in the read. The annotation of the CEN.PK113-7D mitochondrial reference genome is shown at the bottom of each panel.

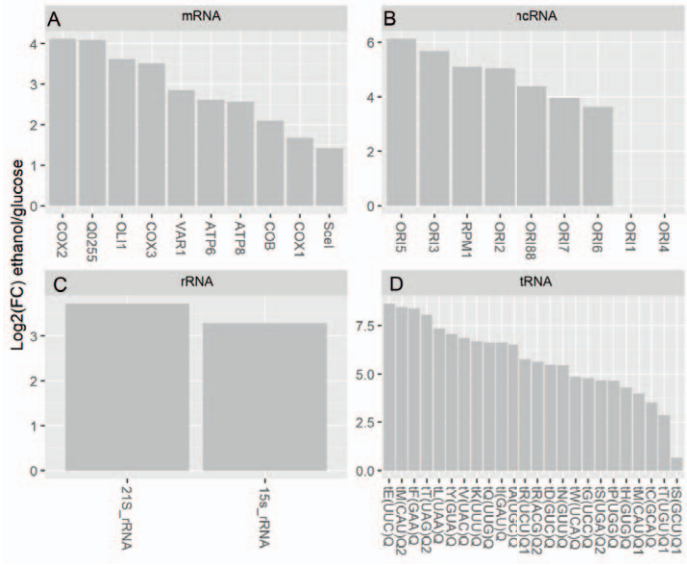


Figure S3.12. Change in expression of mitochondrial RNA species between glucose- and ethanol-grown cultures. Data represent the *log2* fold-change (FC) of transcript expression of mitochondrial RNA in ethanol as compared to glucose cultures. A) mRNA (no separation made between introns and exon reads), B) ncRNA, C) rRNA, D) tRNA.

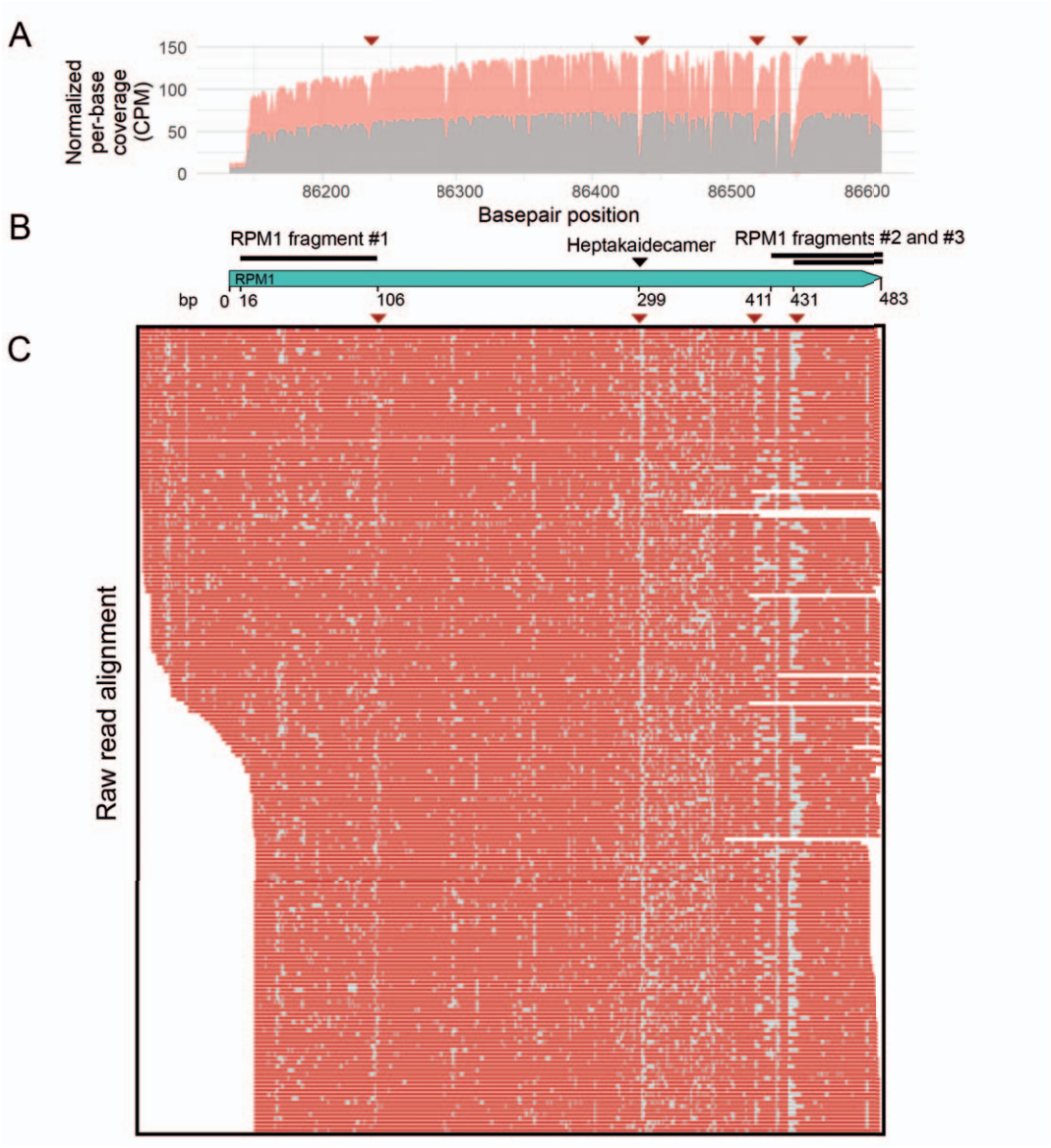


Figure S3.13. Coverage plot and read alignment at the *RPM1* locus. A) Average coverage of triplicate experiments of the gene normalized to CPM on ethanol (grey). The standard deviation of coverage is shown in red. Decreases in coverage caused by RNA processing are indicated with red arrows. B) Schematic representation of the *RPM1* locus, including the RNA fragmentation pattern (black boxes) and heptakaidecamer location (black arrow) proposed by Turk *et al.* [13]. C) Visualization of the sequencing reads mapped to the *RPM1* locus using Tablet. Red lines indicate correct mapping of sequencing reads to the location on the mtDNA, grey lines indicate a gap in the reads. Gaps potentially resulting from RNA processing are indicated with red arrows.

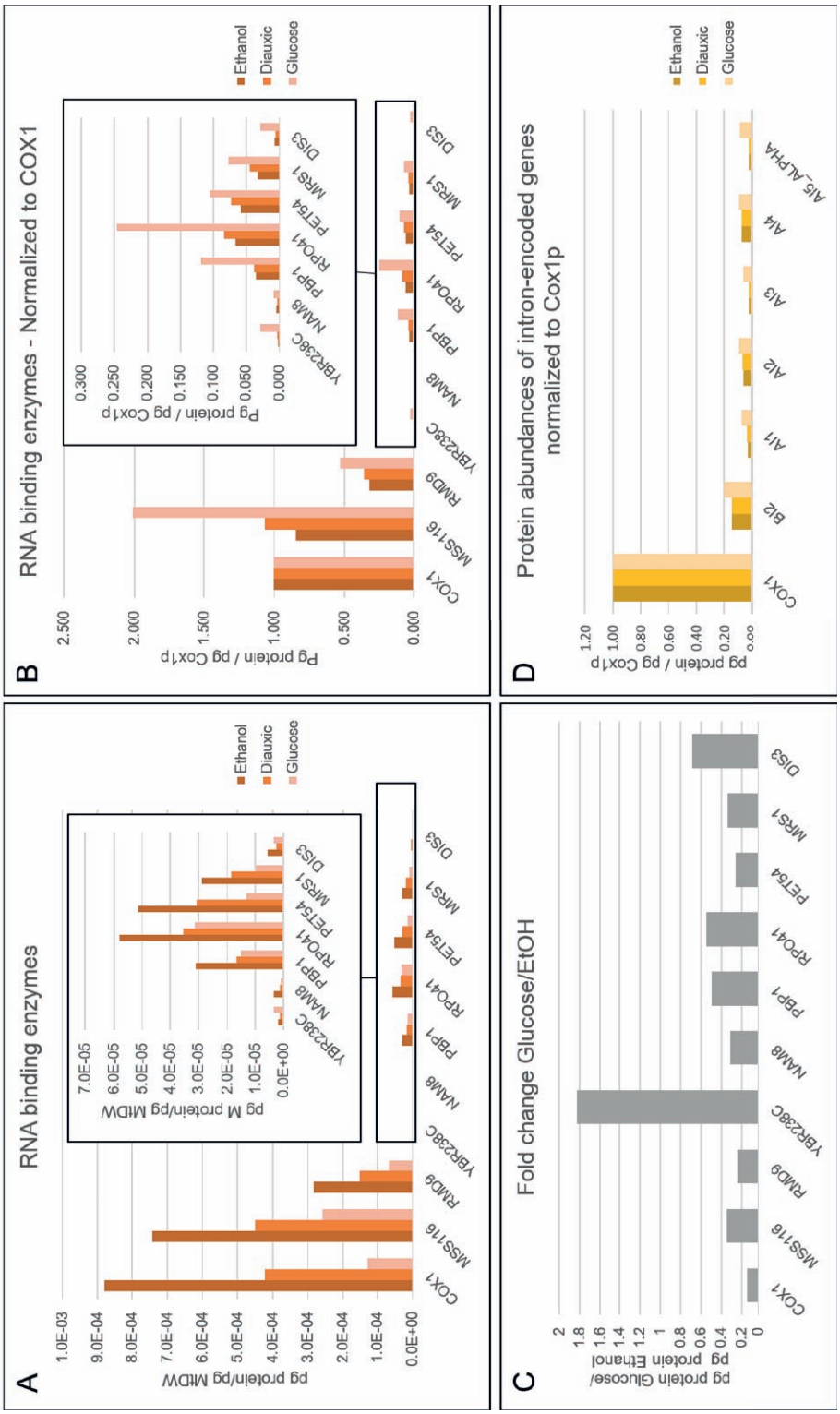


Figure S3.14. Abundance of (putative) mitochondrial RNA binding proteins as determined by Di Bartolomeo *et al.* [47], and the abundance of intron-encoded proteins. A-C) From the datasets from Mitchell *et al* [113] Tsvetanova *et al* [114] of RNA-binding proteins in yeast, a list of RNA-binding proteins was extracted. This list was cross-referenced with the mitochondrial proteome a published by Di Bartolomeo *et al.* to identify a list of nine (putative) mitochondrial RNA binding proteins. All panels in this figure represent protein abundance as reported by di Bartolomeo and colleagues [47]. A) Absolute level of the putative RNA-binding proteins and Cox1p in pg mitochondrial protein normalized by mitochondrial dry weight during growth on ethanol and on glucose and during the diauxic shift. B) Relative abundance of the nine putative RNA binding proteins normalized to the level of Cox1p. C) Fold-change in RNA binding proteins abundance between cultures with glucose and ethanol as carbon source. D) Abundance normalized to Cox1p of intron-encoded proteins during growth on ethanol and on glucose and during the diauxic shift.

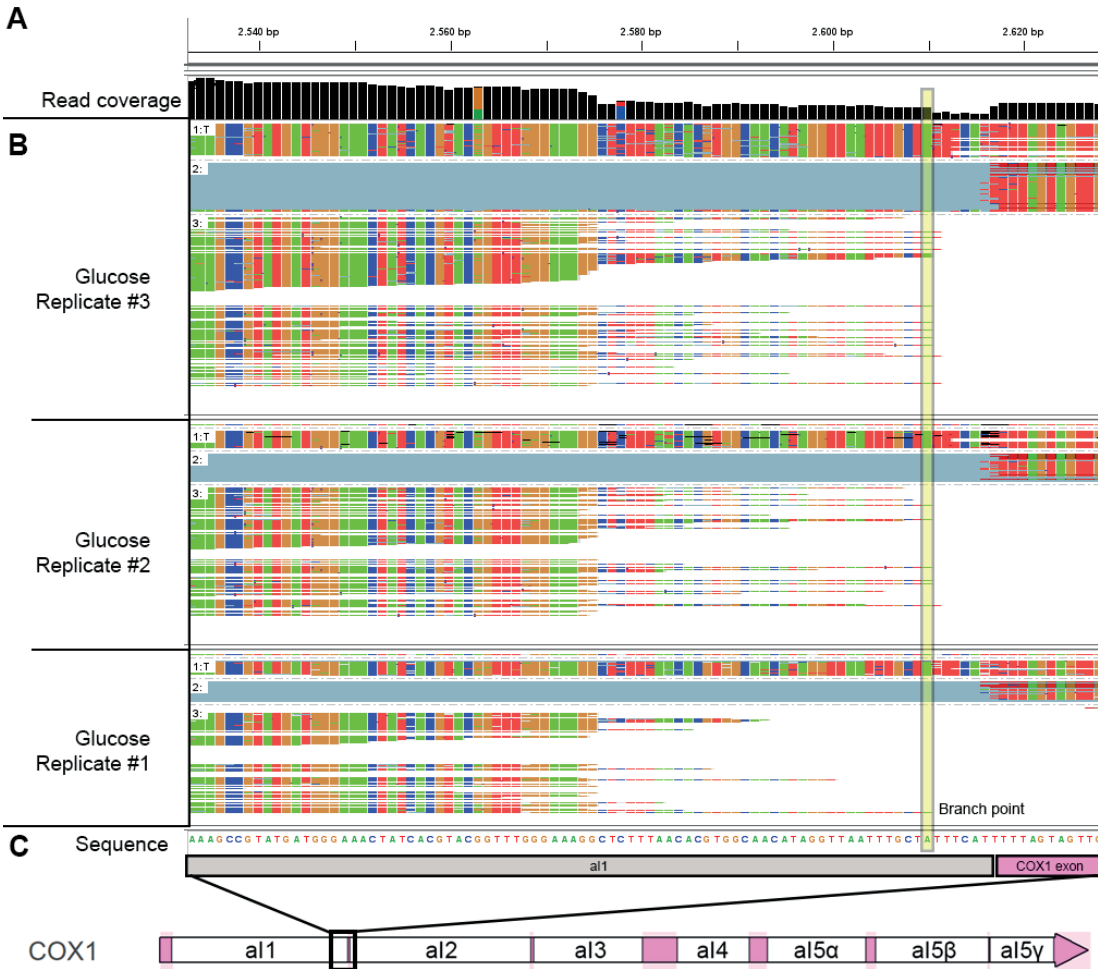


Figure S3.15. Read mapping at the 3' branch point of *COX1 al1* intron sequences. A) Coverage plot (in black) of the *al1* locus. B) Mapping of raw reads for three separate replicates mapped to the reference sequence. Each line is one read, colors indicate the different bases of the read (A = green, T = red, C = dark blue, G = orange). Grey-blue lines indicate gaps in the reads. In A) and B) the branch point as described by Yang and colleagues (where the circularization of the RNA through a covalent bond occurs, [95]) is indicated with a yellow box. C) Schematic overview of the DNA sequence of the *COX1-al1* intron-exon junction.

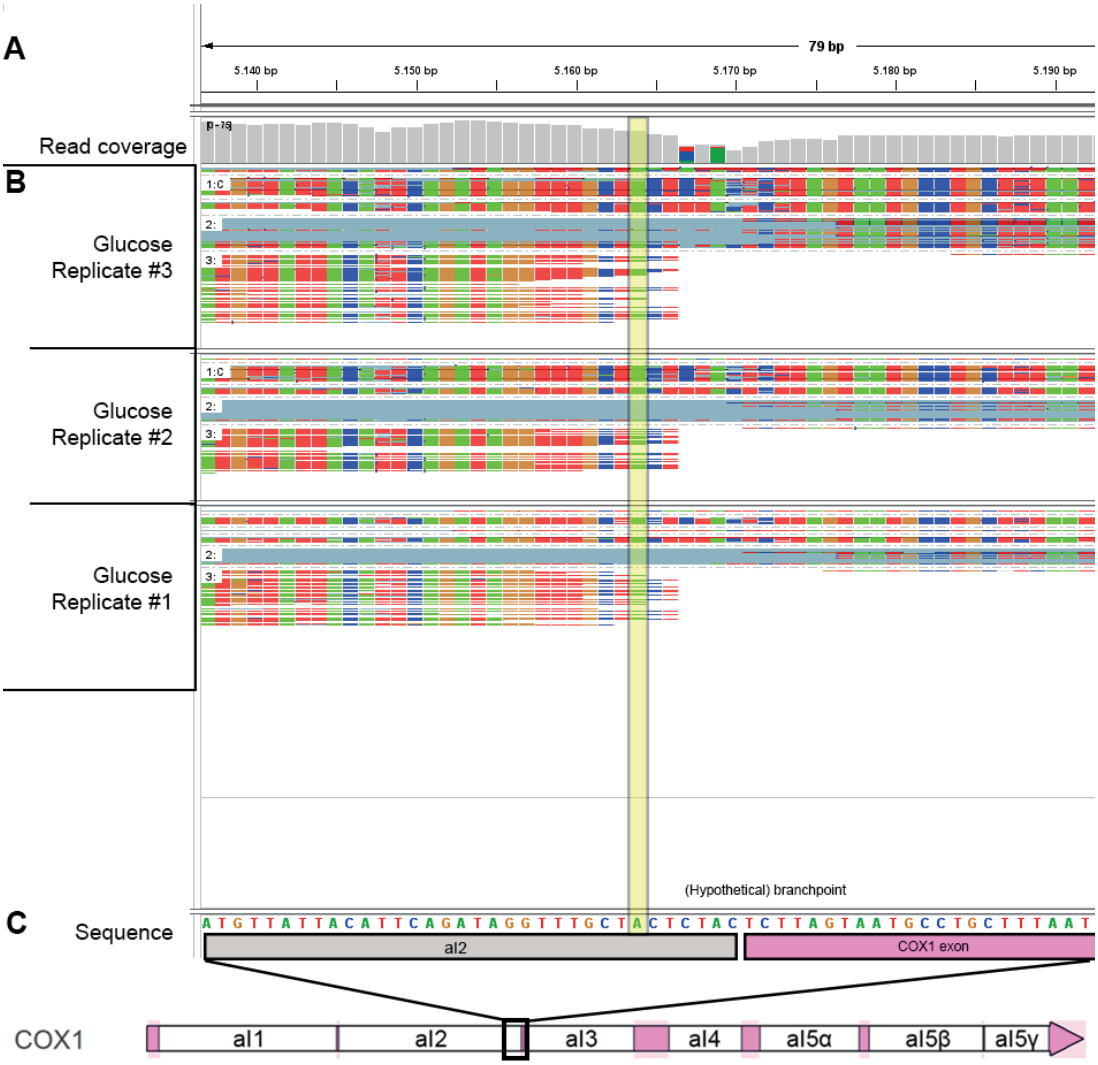


Figure S3.16. Read mapping at the 3' branch point of *COX1 al2* intron sequence. A) Coverage plot (in grey) of the *al2* locus. B) Mapping of raw reads for three separate replicates mapped to the reference sequence. Each line is one read, colors indicate the different bases of the read (A = green, T = red, C = dark blue, G = orange). Grey-blue lines indicate gaps in the reads. The exact branch point of al2 is not described in literature, therefore the branchpoint was hypothesized to be the uracil-residue within a palindromic sequence where reads end, as indicated with a yellow box in A) and B). C) Schematic overview of the DNA sequence of the *COX1-al2* intron-exon junction.

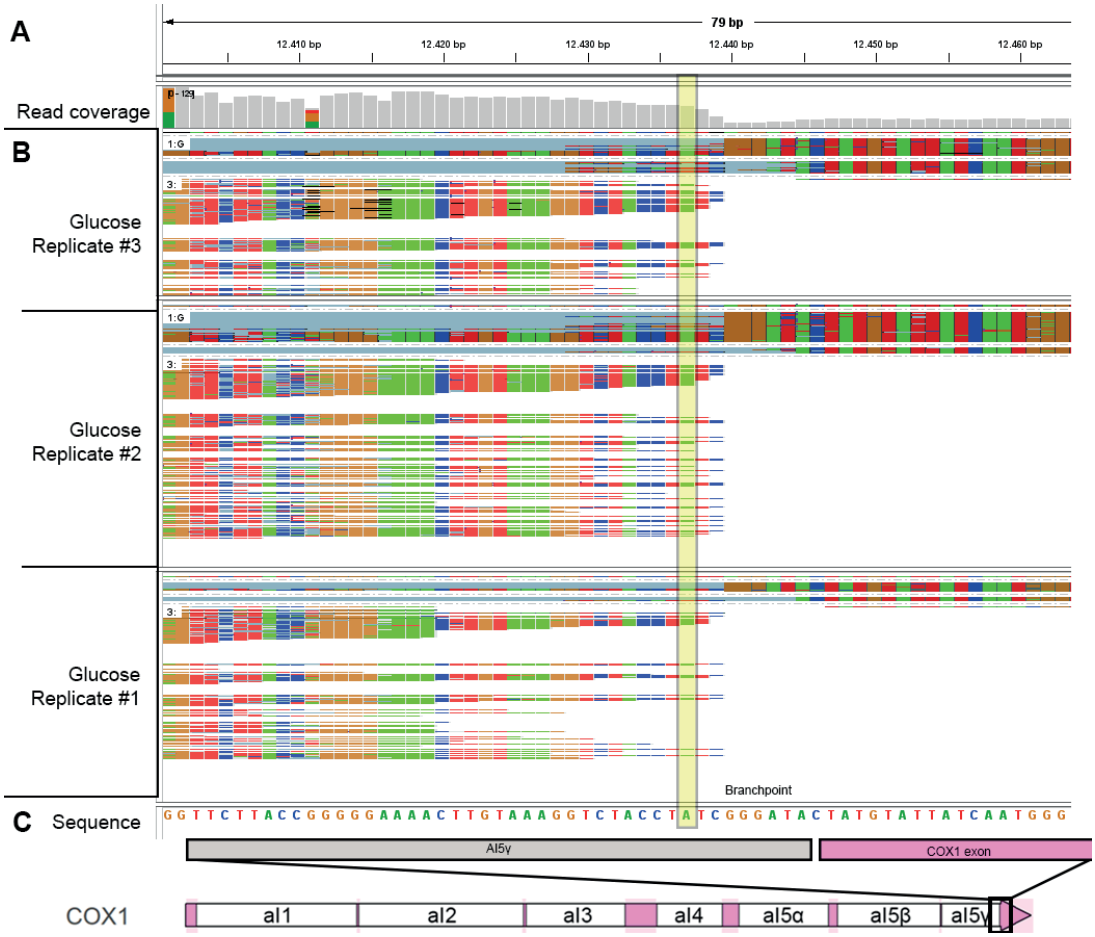


Figure S3.17. Read mapping at the 3' branch point of *COX1 al5γ* intron sequence. A) Coverage plot (in gray) of the *al5γ* locus. B) Mapping of raw reads for three separate replicates mapped to the reference sequence. Each line is one read, colors indicate the different bases of the read (A = green, T = red, C = dark blue, G = orange). Grey-blue lines indicate gaps in the reads. In A) and B) the branch point as described by Schmelzer and Schweyen (where the circularization of the RNA through a covalent bond occurs, [94]) is indicated with a yellow box. C) Schematic overview of the DNA sequence of the *COX1-al5γ* intron-exon junction.

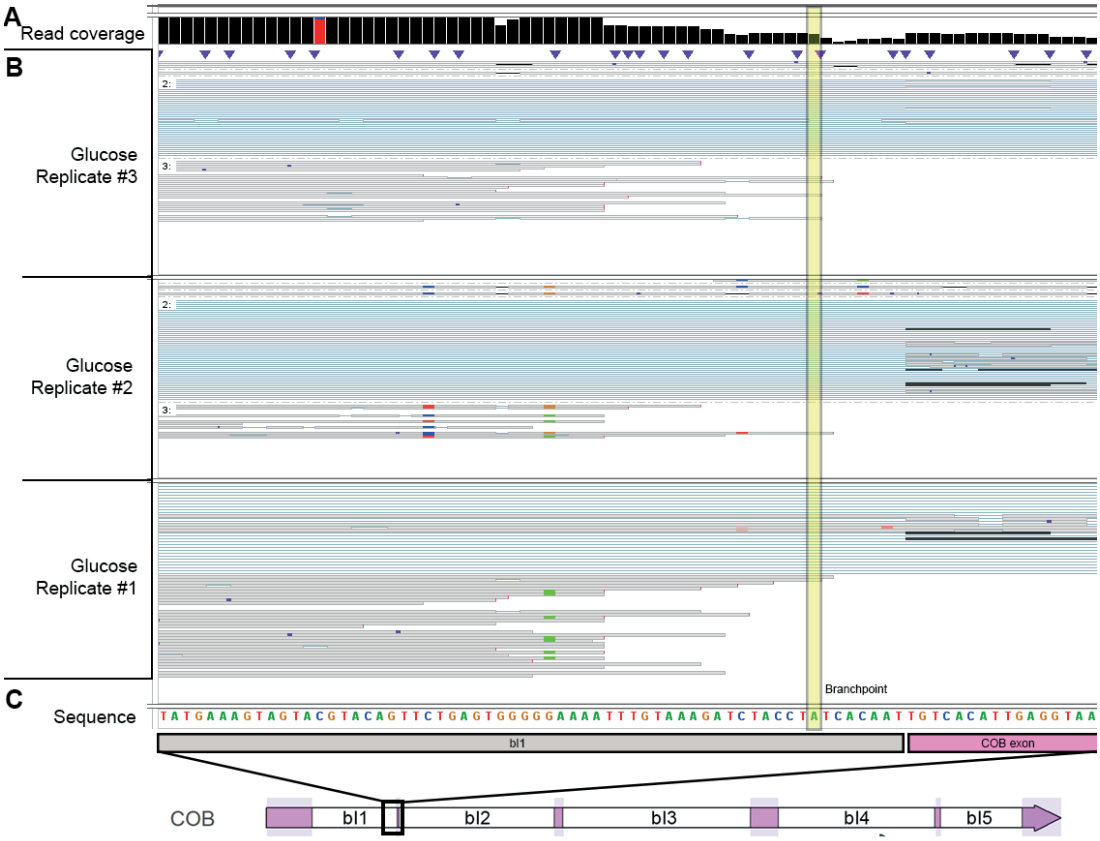


Figure S3.18. Read mapping at the 3' branch point of *COB bI1* intron sequence. A) Coverage plot (in black) of the *bI1* locus. B) Mapping of raw reads for three separate replicates mapped to the reference sequence. Each grey line is one read; blue lines indicate gaps in the reads. In A) and B) the branch point as described by Schmelzer and Schweyen (where the circularization of the RNA through a covalent bond occurs, [94]) is indicated with a yellow box. C) Schematic overview of the DNA sequence of the *COB-bI1* intron-exon junction.

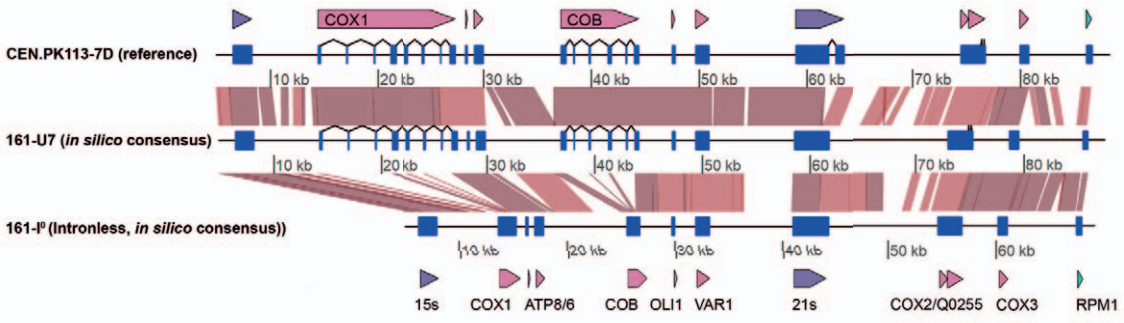


Figure S3.19. Alignment of the mitochondrial genome of CEN.PK113-7D, and consensus sequences of the mitochondrial genomes of 161-U7 and 161-I⁰. The mitochondrial genome is represented as a black line and the respective coordinates are indicated below each genome/contig. The locations of exon mRNA, rRNA and ncRNA are indicated as blue squares, tRNA and *ori* are not shown. Spliced exons are connected by a line, annotations of the sequences are indicated above (CEN.PK113-7D, 161-U7) and below (161-I⁰) the alignment. Red boxes or lines indicate a sequence identity of > 95 % (BLAST) between the two alignments.

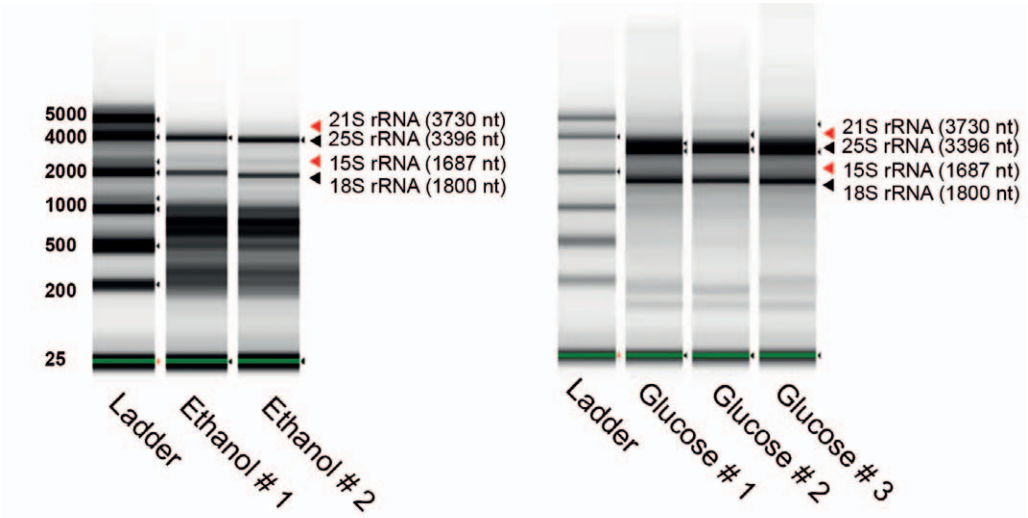


Figure S3.20. Electrophoretic separation of poly(A)-tailed RNA used for Nanopore RNA-seq as determined by a ScreenTape assay. Ladder: RNA ScreenTape ladder. The sizes of major ribosomal RNAs are indicated, along with the expected height, as mitochondrial rRNA run at a different height than expected as described by Locker *et al.* [118].

Supplementary tables S3.1 - S3.8, Supplementary methods, sequences used in this study, and supplementary file SI_readcounts.xlsx can be accessed electronically at https://doi.org/10.1101/2023.01.S9.524680 or via the QR-code below.

Links to: https://doi.org/10.1101/2023.01.S9.524680

References

1. Malina, C., Larsson, C., and Nielsen, J., Yeast mitochondria: an overview of mitochondrial biology and the potential of mitochondrial systems biology. *FEMS Yeast Research*, *2018*. ***18****(5)*.
2. Braun, R.J. and Westermann, B., Mitochondrial dynamics in yeast cell death and aging. *Biochemical Society Transactions*, *2011*. ***39****(5): p. 1520-1526*.
3. Timmis, J.N., Ayliffe, M.A., Huang, C.Y., and Martin, W., Endosymbiotic gene transfer: organelle genomes forge eukaryotic chromosomes. *Nature Reviews Genetics*, *2004*. ***5****(2): p. 123-135*.
4. Contamine, V. and Picard, M., Maintenance and integrity of the mitochondrial genome: A plethora of nuclear genes in the budding yeast. *Microbiology and Molecular Biology Reviews*, *2000*. ***64****(2): p. 281-315*.
5. Baile, M.G. and Claypool, S.M., The power of yeast to model diseases of the powerhouse of the cell. *Frontiers in bioscience (Landmark edition)*, *2013*. ***18****: p. 241*.
6. Larosa, V. and Remacle, C., Transformation of the mitochondrial genome. *International Journal of Developmental Biology*, *2013*. ***57****(6-7-8): p. 659-665*.
7. Yoon, Y.G. and Koob, M.D., Intramitochondrial transfer and engineering of mammalian mitochondrial genomes in yeast. *Mitochondrion*, *2019*. ***46****: p. 15-21*.
8. Foury, F., Roganti, T., Lecrenier, N., and Purnelle, B., The complete sequence of the mitochondrial genome of *Saccharomyces cerevisiae*. *FEBS letters*, *1998*. ***440****(3): p. 325-331*.
9. Basu, U., Bostwick, A.M., Das, K., Dittenhafer-Reed, K.E., and Patel, S.S., Structure, mechanism, and regulation of mitochondrial DNA transcription initiation. *Journal of Biological Chemistry*, *2020*. ***295****(52): p. 18406-18425*.
10. Grivell, L.A., Nucleo-mitochondrial interactions in yeast mitochondrial biogenesis. *European Journal of Biochemistry*, *1989*. ***182****(3): p. 477-93*.
11. Ojala, D., Montoya, J., and Attardi, G., tRNA punctuation model of RNA processing in human mitochondria. *Nature*, *1981*. ***290****(5806): p. 470-474*.
12. Osinga, K.A., De Vries, E., Van der Horst, G., and Tabak, H.F., Processing of yeast mitochondrial messenger RNAs at a conserved dodecamer sequence. *The EMBO journal*, *1984*. ***3****(4): p. 829-834*.
13. Turk, E.M., Das, V., Seibert, R.D., and Andrulis, E.D., The mitochondrial RNA landscape of *Saccharomyces cerevisiae*. *PLoS One*, *2013*. ***8****(10): p. e78105*.
14. Schreiber, K., Csaba, G., Haslbeck, M., and Zimmer, R., Alternative splicing in next generation sequencing data of *Saccharomyces cerevisiae*. *PLoS One*, *2015*. ***10****(10): p. e0140487*.
15. Pel, H.J. and Grivell, L.A., The biology of yeast mitochondrial introns. *Molecular Biology Reports*, *1993*. ***18****(1): p. 1-13*.
16. Michel, F., A maturase-like coding sequence downstream of the *OXI2* gene of yeast mitochondrial DNA is interrupted by two GC clusters and a putative end-of-messenger signal. *Current genetics*, *1984*. ***8****(4): p. 307-317*.
17. Bordonné, R., Dirheimer, G., and Martin, R.P., Expression of the *oxi1* and maturase-related RF1 genes in yeast mitochondria. *Current genetics*, *1988*. ***13****: p. 227-233*.
18. Colleaux, L., d'Auriol, L., Betermier, M., Cottarel, G., Jacquier, A., Galibert, F., and Dujon, B., Universal code equivalent of a yeast mitochondrial intron reading frame is expressed into *E. coli* as a specific double strand endonuclease. *Cell*, *1986*. ***44****(4): p. 521-33*.
19. Lipinski, K.A., Kaniak-Golik, A., and Golik, P., Maintenance and expression of the *S. cerevisiae* mitochondrial genome—From genetics to evolution and systems biology. *Biochimica et Biophysica Acta (BBA) - Bioenergetics*, *2010*. ***1797****(6): p. 1086-1098*.
20. Lambowitz, A.M. and Belfort, M., Introns as mobile genetic elements. *Annual review of biochemistry*, *1993*. ***62****: p. 587-622*.
21. Nookaew, I., Papini, M., Pornputtapong, N., Scalcinati, G., Fagerberg, L., Uhlén, M., and Nielsen, J., A comprehensive comparison of RNA-Seq-based transcriptome analysis from reads to differential gene expression and cross-comparison with microarrays: a case study in *Saccharomyces cerevisiae*. *Nucleic acids research*, *2012*. ***40****(20): p. 10084-97*.
22. Jenjaroenpun, P., Wongsurawat, T., Pereira, R., Patumcharoenpol, P., Ussery, D.W., Nielsen, J., and Nookaew, I., Complete genomic and transcriptional landscape analysis using third-generation sequencing: a case study of *Saccharomyces cerevisiae* CEN. PK113-7D. *Nucleic acids research*, *2018*. ***46****(7): p. e38-e38*.
23. Gagliardi, D., Stepien, P.P., Temperley, R.J., Lightowlers, R.N., and Chrzanowska-Lightowlers, Z.M.A., Messenger RNA stability in mitochondria: different means to an end. *Trends in Genetics*, *2004*. ***20****(6): p. 260-267*.
24. Osorio, D. and Cai, J.J., Systematic determination of the mitochondrial proportion in human and mice tissues for single-cell RNA-sequencing data quality control. *Bioinformatics*, *2021*. ***37****(7): p. 963-967*.
25. Mueller, D.M. and Getz, G.S., Steady state analysis of mitochondrial RNA after growth of yeast *Saccharomyces cerevisiae* under catabolite repression and derepression. *Journal of Biological Chemistry*, *1986*. ***261****(25): p. 11816-11822*.
26. Depledge, D.P., Srinivas, K.P., Sadaoka, T., Bready, D., Mori, Y., Placantonakis, D.G., Mohr, I., and Wilson, A.C., Direct RNA sequencing on nanopore arrays redefines the transcriptional complexity of a viral pathogen. *Nature Communications*, *2019*. ***10****(1): p. 754*.
27. Roach, N.P., Sadowski, N., Alessi, A.F., Timp, W., Taylor, J., and Kim, J.K., The full-length transcriptome of *C. elegans* using direct RNA sequencing. *Genome research*, *2020*. ***30****(2): p. 299-312*.
28. Zhao, L., Zhang, H., Kohnen, M.V., Prasad, K.V.S.K., Gu, L., and Reddy, A.S.N., Analysis of transcriptome and epitranscriptome in plants using PacBio Iso-Seq and Nanopore-based direct RNA sequencing. *Frontiers in Genetics*, *2019*. ***10****: p. 253*.
29. Workman, R.E., Tang, A.D., Tang, P.S., Jain, M., Tyson, J.R., Razaghi, R., Zuzarte, P.C., Gilpatrick, T., Payne, A., and Quick, J., Nanopore native RNA sequencing of a human poly (A) transcriptome. *Nature methods*, *2019*. ***16****(12): p. 1297-1305*.
30. Ono, H. and Yoshida, M.-a., Direct RNA sequencing approach to compare non-model mitochondrial transcriptomes: An application to a cephalopod host and its mesozoan parasite. *Methods*, *2020*. ***176****: p. 55-61*.
31. Gao, Y., Liu, X., Wu, B., Wang, H., Xi, F., Kohnen, M.V., Reddy, A.S.N., and Gu, L., Quantitative profiling of N6-methyladenosine at single-base resolution in stem-differentiating xylem of *Populus trichocarpa* using Nanopore direct RNA sequencing. *Genome Biology*, *2021*. ***22****(1): p. 22*.
32. Entian, K.-D. and Kötter, P., 25 yeast genetic strain and plasmid collections. *Methods in Microbiology*, *2007*.
33. Mahler, H.R. and Lin, C.C., Molecular events during the release of delta-aminolevulinate dehydratase from catabolite repression. *Journal of Bacteriology*, *1978*. ***135****(1): p. 54-61*.
34. Wenzlau, J.M., Saldanha, R.J., Butow, R.A., and Perlman, P.S., A latent intron-encoded maturase is also an endonuclease needed for intron mobility. *Cell*, *1989*. ***56****(3): p. 421-430*.
35. Huang, H.-R., Rowe, C.E., Mohr, S., Jiang, Y., Lambowitz, A.M., and Perlman, P.S., The splicing of yeast mitochondrial group I and group II introns requires a DEAD-box protein with RNA chaperone function. *Proceedings of the National Academy of Sciences*, *2005*. ***102****(1): p. 163-168*.
36. Verduyn, C., Postma, E., Scheffers, W.A., and Van Dijken, J.P., Effect of benzoic acid on metabolic fluxes in yeasts: A

continuous-culture study on the regulation of respiration and alcoholic fermentation. *Yeast*, *1992.* ***8****(7): p. 501-517.*

37. Pronk, J.T., Auxotrophic yeast strains in fundamental and applied research. *Applied and environmental microbiology*, *2002.* ***68****(5): p. 2095-2100.*
38. Botman, D., de Groot, D.H., Schmidt, P., Goedhart, J., and Teusink, B., *In vivo* characterisation of fluorescent proteins in budding yeast. *Scientific reports*, *2019.* ***9****(1): p. 2234.*
39. Lee, M.E., DeLoache, W.C., Cervantes, B., and Dueber, J.E., A highly characterized yeast toolkit for modular, multipart assembly. *ACS synthetic biology*, *2015.* ***4****(9): p. 975-986.*
40. Bouwknegt, J., Koster, C.C., Vos, A.M., Ortiz-Merino, R.A., Wassink, M., Luttik, M.A.H., van den Broek, M., Hagedoorn, P.L., and Pronk, J.T., Class-II dihydroorotate dehydrogenases from three phylogenetically distant fungi support anaerobic pyrimidine biosynthesis. *Fungal Biology and Biotechnology*, *2021.* ***8****(1): p. 10.*
41. Boonekamp, F.J., Knibbe, E., Vieira-Lara, M.A., Wijsman, M., Luttik, M.A.H., van Eunen, K., Ridder, M.d., Bron, R., Almonacid Suarez, A.M., van Rijn, P.,... Daran-Lapujade, P., Full humanization of the glycolytic pathway in *Saccharomyces cerevisiae*. *Cell reports*, *2022.* ***39****(13): p. 111010.*
42. Gietz, R.D. and Woods, R.A., Transformation of yeast by lithium acetate/single-stranded carrier DNA/polyethylene glycol method, in Methods in enzymology. 2002, *Elsevier*. p. 87-96.
43. Randazzo, P., Bennis, N.X., Daran, J.-M., and Daran-Lapujade, P., gEL DNA: A cloning- and polymerase chain reaction–free method for CRISPR-based multiplexed genome editing. *The CRISPR Journal*, *2021.* ***4****(6): p. 896-913.*
44. Gorter de Vries, A.R., de Groot, P.A., van den Broek, M., and Daran, J.-M.G., CRISPR-Cas9 mediated gene deletions in lager yeast *Saccharomyces pastorianus*. *Microbial cell factories*, *2017.* ***16****(1): p. 222.*
45. Liao, P.-C., Boldogh, I.R., Siegmund, S.E., Freyberg, Z., and Pon, L.A., Isolation of mitochondria from *Saccharomyces cerevisiae* using magnetic bead affinity purification. *PloS one*, *2018.* ***13****(4): p. e0196632.*
46. Luttik, M.A., Overkamp, K.M., Kötter, P., De Vries, S., Van Dijken, J.P., and Pronk, J.T., The *Saccharomyces cerevisiae NDE1* and *NDE2* genes encode separate mitochondrial NADH dehydrogenases catalyzing the oxidation of cytosolic NADH. *Journal of Biological Chemistry*, *1998.* ***273****(38): p. 24529-24534.*
47. Di Bartolomeo, F., Malina, C., Campbell, K., Mormino, M., Fuchs, J., Vorontsov, E., Gustafsson, C.M., and Nielsen, J., Absolute yeast mitochondrial proteome quantification reveals trade-off between biosynthesis and energy generation during diauxic shift. *Proceedings of the National Academy of Sciences*, *2020.* ***117****(13): p. 7524-7535.*
48. Doerr, A., De Reus, E., Van Nies, P., Van der Haar, M., Wei, K., Kattan, J., Wahl, A., and Danelon, C., Modelling cell-free RNA and protein synthesis with minimal systems. *Physical biology*, *2019.* ***16****(2): p. 025001.*
49. Nijkamp, J., van den Broek, M., Datema, E., de Kok, S., Bosman, L., Luttik, M., Pronk, J., de Ridder, D., and Daran, J.-M., *De novo* sequencing, assembly and analysis of the genome of the laboratory strain *Saccharomyces cerevisiae* CEN. PK113-7D, a model for modern industrial biotechnology. *Microbial cell factories*, *2012.* ***11****: p. 36.*
50. Salazar, A.N., Gorter de Vries, A.R., van den Broek, M., Wijsman, M., de la Torre Cortés, P., Brickwedde, A., Brouwers, N., Daran, J.-M.G., and Abeel, T., Nanopore sequencing enables near-complete *de novo* assembly of *Saccharomyces cerevisiae* reference strain CEN.PK113-7D. *FEMS Yeast Research*, *2017.* ***17****(7).*
51. Cherry, J.M., Hong, E.L., Amundsen, C., Balakrishnan, R., Binkley, G., Chan, E.T., Christie, K.R., Costanzo, M.C., Dwight, S.S., Engel, S.R.,... Wong, E.D., *Saccharomyces* Genome Database: the genomics resource of budding yeast. *Nucleic acids research*, *2012.* ***40****(Database issue): p. D700-5.*
52. Cherry, J.M., Ball, C., Weng, S., Juvik, G., Schmidt, R., Adler, C., Dunn, B., Dwight, S., Riles, L., Mortimer, R.K., and Botstein, D., Genetic and physical maps of *Saccharomyces cerevisiae*. *Nature*, *1997.* ***387****(6632 Suppl): p. 67-73.*
53. Gu, Z., Gu, L., Eils, R., Schlesner, M., and Brors, B., Circlize implements and enhances circular visualization in R. *Bioinformatics*, *2014.* ***30****(19): p. 2811-2812.*
54. Li, H., Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics*, *2018.* ***34****(18): p. 3094-3100.*
55. Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., and Durbin, R., The Sequence Alignment/Map format and SAMtools. *Bioinformatics*, *2009.* ***25****(16): p. 2078-9.*
56. Liao, Y., Smyth, G.K., and Shi, W., featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics*, *2013.* ***30****(7): p. 923-930.*
57. Robinson, M.D., McCarthy, D.J., and Smyth, G.K., edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics*, *2009.* ***26****(1): p. 139-140.*
58. Robinson, J.T., Thorvaldsdóttir, H., Winckler, W., Guttman, M., Lander, E.S., Getz, G., and Mesirov, J.P., Integrative genomics viewer. *Nature biotechnology*, *2011.* ***29****(1): p. 24-6.*
59. Milne, I., Stephen, G., Bayer, M., Cock, P.J.A., Pritchard, L., Cardle, L., Shaw, P.D., and Marshall, D., Using Tablet for visual exploration of second-generation sequencing data. *Briefings in Bioinformatics*, *2012.* ***14****(2): p. 193-202.*
60. Schindelin, J., Arganda-Carreras, I., Frise, E., Kaynig, V., Longair, M., Pietzsch, T., Preibisch, S., Rueden, C., Saalfeld, S., Schmid, B.,... Cardona, A., Fiji: an open-source platform for biological-image analysis. *Nature methods*, *2012.* ***9****(7): p. 676-682.*
61. Li, H. and Durbin, R., Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics*, *2009.* ***25****(14): p. 1754-60.*
62. Bankevich, A., Nurk, S., Antipov, D., Gurevich, A., Dvorkin, M., Kulikov, A., Lesin, V., Nikolenko, S., Pham, S., and Prjibelski, A., Pyshkin A. v, Sirotkin A. v, Vyahhi N, Tesler G, Alekseyev MA, Pevzner PA. 2012. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *Journal of Computational Biology*, *2012.* ***19****: p. 455-477.*
63. Kurtz, S., Phillippy, A., Delcher, A.L., Smoot, M., Shumway, M., Antonescu, C., and Salzberg, S.L., Versatile and open software for comparing large genomes. *Genome Biology*, *2004.* ***5****(2): p. 1-9.*
64. Guy, L., Roat Kultima, J., and Andersson, S.G.E., GenoPlotR: comparative gene and genome visualization in R. *Bioinformatics*, *2010.* ***26****(18): p. 2334-2335.*
65. Mercer, T.R., Neph, S., Dinger, M.E., Crawford, J., Smith, M.A., Shearwood, A.-M.J., Haugen, E., Bracken, C.P., Rackham, O., and Stamatoyannopoulos, J.A.J.C., The human mitochondrial transcriptome. *Cell*, *2011.* ***146****(4): p. 645-658.*
66. Simon, M. and Faye, G., Organization and processing of the mitochondrial *oxi3/oli2* multigenic transcript in yeast. *Molecular Genetics and Genomics*, *1984.* ***196****(2): p. 266-74.*
67. Thalenfeld, B.E., Bonitz, S.G., Nobrega, F.G., Macino, G., and Tzagoloff, A., *oli1* transcripts in wild type and in a cytoplasmic "petite" mutant of yeast. *Journal of Biological Chemistry*, *1983.* ***258****(23): p. 14065-8.*
68. Turk, E.M. and Caprara, M.G., Splicing of yeast al5β group I intron requires SUV3 to recycle *MRS1* via mitochondrial degradosome-promoted decay of excised intron ribonucleoprotein (RNP). *Journal of Biological Chemistry*, *2010.* ***285****(12): p. 8585-8594.*
69. Margossian, S.P., Li, H., Zassenhaus, H.P., and Butow, R.A., The DExH box protein Suv3p is a component of a yeast mitochondrial 3'-to-5' exoribonuclease that suppresses group I intron toxicity. *Cell*, *1996.* ***84****(2): p. 199-209.*
70. Wilson, B.D., Eisenstein, M., and Soh, H.T., High-Fidelity Nanopore Sequencing of Ultra-Short DNA Targets. *Analytical Chemistry*, *2019.* ***91****(10): p. 6783-6789.*
71. Hillen, H.S., Markov, D.A., Wojtas, I.D., Hofmann, K.B., Lidschreiber, M., Cowan, A.T., Jones, J.L., Temiakov, D.,

Cramer, P., and Anikin, M., The pentatricopeptide repeat protein Rmd9 recognizes the dodecameric element in the 3'-UTRs of yeast mitochondrial mRNAs. *Proceedings of the National Academy of Sciences*, *2021*. ***118****(15): p. e2009329118*.

72. Nouet, C., Bourens, M., Hlavacek, O., Marsy, S., Lemaire, C., and Dujardin, G., Rmd9p controls the processing/stability of mitochondrial mRNAs and its overexpression compensates for a partial deficiency of oxa1p in *Saccharomyces cerevisiae*. *Genetics*, *2007*. ***175****(3): p. 1105-15*.
73. Szczesny, R.J., Borowski, L.S., Malecki, M., Wojcik, M.A., Stepien, P.P., and Golik, P., RNA Degradation in Yeast and Human Mitochondria. *Biochimica et Biophysica Acta (BBA) - Gene Regulatory Mechanisms*, *2012*. ***1819****(9): p. 1027-1034*.
74. Visser, W., van Spronsen, E.A., Nanninga, N., Pronk, J.T., Kuenen, J.G., and van Dijken, J.P., Effects of growth conditions on mitochondrial morphology in *Saccharomyces cerevisiae*. *Antonie van Leeuwenhoek*, *1995*. ***67****(3): p. 243-253*.
75. Altmann, K., Dürr, M., and Westermann, B., *Saccharomyces cerevisiae* as a model organism to study mitochondrial biology, in Mitochondria. 2007, *Springer*. p. 81-90.
76. Dong, X., Tian, L., Gouil, Q., Kariyawasam, H., Su, S., De Paoli-Iseppi, R., Prawer, Yair David J., Clark, M.B., Breslin, K., Iminitoff, M., Blewitt, M.E., Law, C.W., and Ritchie, M.E., The long and the short of it: unlocking nanopore long-read RNA sequencing data with short-read differential expression analysis tools. *NAR Genomics and Bioinformatics*, *2021*. ***3****(2)*.
77. Soneson, C., Yao, Y., Bratus-Neuenschwander, A., Patrignani, A., Robinson, M.D., and Hussain, S., A comprehensive examination of Nanopore native RNA sequencing for characterization of complex transcriptomes. *Nature Communications*, *2019*. ***10****(1): p. 3359*.
78. Merle, P. and Kadenbach, B., The Subunit Composition of Mammalian Cytochrome *c* Oxidase. *European Journal of Biochemistry*, *1980*. ***105****(3): p. 499-507*.
79. Fontanesi, F., Soto, I.C., and Barrientos, A., Cytochrome *c* oxidase biogenesis: New levels of regulation. *IUBMB Life*, *2008*. ***60****(9): p. 557-568*.
80. Tsukihara, T., Aoyama, H., Yamashita, E., Tomizaki, T., Yamaguchi, H., Shinzawa-Itoh, K., Nakashima, R., Yaono, R., and Yoshikawa, S., The whole structure of the 13-subunit oxidized cytochrome c oxidase at 2.8 Å. *Science*, *1996*. ***272****(5265): p. 1136-1144*.
81. Baldacci, G. and Zennaro, E., Mitochondrial Transcripts in Glucose-Repressed Cells of *Saccharomyces cerevisiae*. *European Journal of Biochemistry*, *1982*. ***127****(2): p. 411-416*.
82. Sulo, P., Groom, K.R., Wise, C., Steffen, M., and Martin, N., Successful transformation of yeast mitochondria with *RPM1*: an approach for *in vivo* studies of mitochondrial RNase P RNA structure, function and biosynthesis. *Nucleic Acids Research*, *1995*. ***23****(5): p. 856-860*.
83. Rossmanith, W., Of P and Z: Mitochondrial tRNA processing enzymes. *Biochimica et Biophysica Acta (BBA) - Gene Regulatory Mechanisms*, *2012*. ***1819****(9): p. 1017-1026*.
84. Bernardi, G., Lessons from a small, dispensable genome: The mitochondrial genome of yeast. *Gene*, *2005*. ***354****: p. 189-200*.
85. Chen, X.J. and Clark-Walker, G.D., Unveiling the mystery of mitochondrial DNA replication in yeasts. *Mitochondrion*, *2018*. ***38****: p. 17-22*.
86. Baldacci, G., Chérif-Zahar, B., and Bernardi, G., The initiation of DNA replication in the mitochondrial genome of yeast. *The EMBO journal*, *1984*. ***3****(9): p. 2115-20*.
87. Graves, T., Dante, M., Eisenhour, L., and Christianson, T.W., Precise mapping and characterization of the RNA primers of DNA replication for a yeast hypersuppressive petite by *in vitro* capping with guanylyltransferase. *Nucleic acids research*, *1998*. ***26****(5): p. 1309-1316*.
88. Kaufman, B.A., Newman, S.M., Hallberg, R.L., Slaughter, C.A., Perlman, P.S., and Butow, R.A., *In organello* formaldehyde crosslinking of proteins to mtDNA: Identification of bifunctional proteins. *Proceedings of the National Academy of Sciences*, *2000*. ***97****(14): p. 7772-7777*.
89. Dujon, B., Mitochondrial genetics revisited. *Yeast*, *2020*. ***37****(2): p. 191-205*.
90. Séraphin, B., Boulet, A., Simon, M., and Faye, G., Construction of a yeast strain devoid of mitochondrial introns and its use to screen nuclear genes involved in mitochondrial splicing. *Proceedings of the National Academy of Sciences*, *1987*. ***84****(19): p. 6810-6814*.
91. Tzagoloff, A. and Myers, A.M., Genetics Of Mitochondrial Biogenesis. *Annual Review of Biochemistry*, *1986*. ***55****(1): p. 249-285*.
92. Moran, J.V., Zimmerly, S., Eskes, R., Kennell, J.C., Lambowitz, A.M., Butow, R.A., and Perlman, P.S., Mobile group II introns of yeast mitochondrial DNA are novel site-specific retroelements. *Molecular and Cellular Biology*, *1995*. ***15****(5): p. 2828-38*.
93. De Silva, D., Poliquin, S., Zeng, R., Zamudio-Ochoa, A., Marrero, N., Perez-Martinez, X., Fontanesi, F., and Barrientos, A., The DEAD-box helicase Mss116 plays distinct roles in mitochondrial ribogenesis and mRNA-specific translation. *Nucleic Acids Research*, *2017*. ***45****(11): p. 6628-6643*.
94. Schmelzer, C. and Schweyen, R.J., Self-splicing of group II introns *in vitro*: Mapping of the branch point and mutational inhibition of lariat formation. *Cell*, *1986*. ***46****(4): p. 557-565*.
95. Yang, J., Mohr, G., Perlman, P.S., and Lambowitz, A.M., Group II intron mobility in yeast mitochondria: target DNA-primed reverse transcription activity of ai1 and reverse splicing into DNA transposition sites *in vitro*. *Journal of molecular biology*, *1998*. ***282****(3): p. 505-523*.
96. Zhao, Q., Liu, J., Deng, H., Ma, R., Liao, J.Y., Liang, H., Hu, J., Li, J., Guo, Z., Cai, J., Xu, X., Gao, Z., and Su, S., Targeting mitochondria-located circRNA SCAR alleviates NASH via reducing mROS output. *Cell*, *2020*. ***183****(1): p. 76-93.e22*.
97. Zhao, Y., Sun, L., Wang, R.R., Hu, J.-F., and Cui, J., The effects of mitochondria-associated long noncoding RNAs in cancer mitochondria: New players in an old arena. *Critical Reviews in Oncology/Hematology*, *2018*. ***131****: p. 76-82*.
98. Cavalcante, G.C., Magalhães, L., Ribeiro-dos-Santos, Â., and Vidal, A.F., Mitochondrial Epigenetics: Non-Coding RNAs as a Novel Layer of Complexity. *International Journal of Molecular Sciences*, *2020*. ***21****(5): p. 1838*.
99. Huang, H.-R., Functional studies of intron-and nuclear-encoded splicing factors in the mitochondria of *Saccharomyces cerevisiae*. 2004, University of Texas Southwestern Medical Center at Dallas.
100. Rudan, M., Dib, P.B., Musa, M., Kanunnikau, M., Sobočanec, S., Rueda, D., Warnecke, T., and Kriško, A., Normal mitochondrial function in *Saccharomyces cerevisiae* has become dependent on inefficient splicing. *Elife*, *2018*. ***7****: p. e35330*.
101. Thomas, N.K., Poodari, V.C., Jain, M., Olsen, H.E., Akeson, M., and Abu-Shumays, R.L., Direct Nanopore sequencing of individual full length tRNA strands. *ACS Nano*, *2021*. ***15****(10): p. 16642-16653*.
102. Boldogh, I.R. and Pon, L.A., Purification and subfractionation of mitochondria from the yeast *Saccharomyces cerevisiae*, in Methods in cell biology. 2007, *Academic Press*. p. 45-64.
103. Wangsanuwat, C., Heom, K.A., Liu, E., O'Malley, M.A., and Dey, S.S., Efficient and cost-effective bacterial mRNA sequencing from low input samples through ribosomal RNA depletion. *BMC Genomics*, *2020*. ***21****(1): p. 717*.
104. Evers, D.L., Fowler, C.B., Cunningham, B.R., Mason, J.T., and O'Leary, T.J., The effect of formaldehyde fixation on RNA: optimization of formaldehyde adduct removal. *The Journal of Molecular Diagnostics*, *2011*. ***13****(3): p. 282-8*.
105. Schwalb, B., Michel, M., Zacher, B., Frühauf, K., Demel, C., Tresch, A., Gagneur, J., and Cramer, P., TT-seq maps the human transient transcriptome. *Science*, *2016*. ***352****(6290): p. 1225-1228*.
106. Price, A., Garhyan, J., and Gibas, C., The impact of RNA secondary structure on read start locations on the Illumina sequencing platform. *PLoS One*, *2017*. ***12****(2): p. e0173023*.

107. Pyle, A.M., Group II Intron Self-Splicing. *Annual Reviews of Biophysics*, *2016*. ***45***: *p. 183-205*.
108. Jarrell, K.A., Peebles, C.L., Dietrich, R.C., Romiti, S.L., and Perlman, P.S., Group II intron self-splicing. Alternative reaction conditions yield novel products. *Journal of Biological Chemistry*, *1988*. ***263****(7): p. 3432-3439*.
109. Murray, H.L., Mikheeva, S., Coljee, V.W., Turczyk, B.M., Donahue, W.F., Bar-Shalom, A., and Jarrell, K.A., Excision of group II introns as circles. *Molecular cell*, *2001*. ***8****(1): p. 201-211*.
110. Khalid, M.F., Damha, M.J., Shuman, S., and Schwer, B., Structure-function analysis of yeast RNA debranching enzyme (Dbr1), a manganese-dependent phosphodiesterase. *Nucleic acids research*, *2005*. ***33****(19): p. 6349-60*.
111. Thumuluri, V., Almagro Armenteros, J.J., Johansen, Alexander R., Nielsen, H., and Winther, O., DeepLoc 2.0: multi-label subcellular localization prediction using protein language models. *Nucleic Acids Research*, *2022*. ***50****(W1): p. W228-W234*.
112. Sickmann, A., Reinders, J., Wagner, Y., Joppich, C., Zahedi, R., Meyer, H.E., Schönfisch, B., Perschil, I., Chacinska, A., Guiard, B., Rehling, P., Pfanner, N., and Meisinger, C., The proteome of *Saccharomyces cerevisiae* mitochondria. *Proceedings of the National Academy of Sciences*, *2003*. ***100****(23): p. 13207-13212*.
113. Mitchell, S.F., Jain, S., She, M., and Parker, R., Global analysis of yeast mRNPs. *Nature Structural & Molecular Biology*, *2013*. ***20****(1): p. 127-133*.
114. Tsvetanova, N.G., Klass, D.M., Salzman, J., and Brown, P.O., Proteome-Wide Search Reveals Unexpected RNA-Binding Proteins in *Saccharomyces cerevisiae*. *PloS one*, *2010*. ***5****(9): p. e12671*.
115. Turk, E.M. and Caprara, M.G., Splicing of yeast *ai5β* group I intron requires *SUV3* to recycle *MRS1* via mitochondrial degradosome-promoted decay of excised intron ribonucleoprotein (RNP). *Journal of Biological Chemistry*, *2010*. ***285****(12): p. 8585-8594*.
116. Wolters, J.F., Functional variation in the mitochondrial genome of the yeast *Saccharomyces cerevisiae*. 2018, Binghamton University - State University of New York.
117. Hinkle, P.C., P/O ratios of mitochondrial oxidative phosphorylation. *Biochimica et Biophysica Acta (BBA) - Bioenergetics*, *2005*. ***1706****(1): p. 1-11*.
118. Locker, J., Morimoto, R., Synenki, R.M., and Rabinowitz, M., Analysis of mitochondrial RNA in *Saccharomyces cerevisiae*. *Current genetics*, *1980*. ***1****(2): p. 163-72*.
119. Mans, R., Wijsman, M., Daran-Lapujade, P., and Daran, J.-M., A protocol for introduction of multiple genetic modifications in *Saccharomyces cerevisiae* using CRISPR/Cas9. *FEMS yeast research*, *2018*. ***18****(7): p. foy063*.
120. Postma, E., Verduyn, C., Scheffers, W.A., and Van Dijken, J.P., Enzymic analysis of the crabtree effect in glucose-limited chemostat cultures of *Saccharomyces cerevisiae*. *Applied and environmental microbiology*, *1989*. ***55****(2): p. 468-477*.
121. van Hoek, P., van Dijken, J.P., and Pronk, J.T., Regulation of fermentative capacity and levels of glycolytic enzymes in chemostat cultures of *Saccharomyces cerevisiae*. *Enzyme and microbial technology*, *2000*. ***26****(9-10): p. 724-736*.
122. Bruinenberg, P.M., van Dijken, J.P., Kuenen, J.G., and Scheffers, W.A., Critical parameters in the isolation of mitochondria from *Candida utilis*. *Microbiology*, *1985*. ***131****(5): p. 1035-1042*.
123. Sepuri, N.B.V., Gorla, M., and King, M.P., Mitochondrial Lysyl-tRNA synthetase independent import of tRNA lysine into yeast mitochondria. *PloS one*, *2012*. ***7****(4): p. e35321*.

CHAPTER 4

ARGININE AUXOTROPHY CAUSED BY DELETION OF THE MITOCHONDRIAL SELECTABLE MARKER *ARG8* IN *SACCHAROMYCES CEREVISIAE* CAN BE REVERTED BY LOSS-OF-FUNCTION MUTATIONS IN *UME6*

Charlotte C. Koster, Pascale Daran-Lapujade

Abstract

Mitochondria are essential organelles housing key catabolic and anabolic processes. The yeast *Saccharomyces cerevisiae* is a popular model organism for studying mitochondrial biology, as the robustness of *S. cerevisiae* to mitochondrial mutations makes it one of the few organisms in which mitochondrial DNA can be edited. Engineering of mitochondrial DNA requires mitochondrial selectable marker genes, such as the widely used *ARG8*, encoding acetylornithine aminotransferase, which catalyses an essential step in arginine biosynthesis.
This study revealed a frequently occurring revertant mutation that bypasses arginine auxotrophy in *Δarg8* mutants, hampering the use of *ARG8* as selectable marker. Arginine prototrophic *Δarg8* strains were respiratory-deficient and displayed growth defects in the presence and absence of arginine. Whole genome sequencing identified mutations in *UME6*, a pleiotropic transcriptional regulator responsive to nutritional cues, as the most likely cause for the observed phenotype. The loss of function of Ume6 resulted in derepression of *CAR2*, encoding ornithine transaminase, which was found to be the major cause for *Δarg8* mutants recovery of arginine prototrophy. Double deletion of *ARG8* and *CAR2* resulted in strains with wild-type phenotype in the presence of arginine, and stable arginine auxotrophy. *Δarg8 Δcar2* strains offer a robust alternative to *Δarg8* mutants for mitochondrial targeting and expression studies.

4.1 Introduction

Mitochondria are essential organelles in most eukaryotic species. Besides their role in energy metabolism, mitochondria host many essential biosynthetic pathways, including biosynthesis of iron-sulfur clusters, amino acids, heme, and lipids, and play an important role in cellular aging and programmed cell death [1, 2]. *Saccharomyces cerevisiae* is a popular model organism for mitochondrial biology. Mitochondria are largely conserved among eukaryotes, but unlike e.g., mammalian cells, *S. cerevisiae* is able to generate ATP without active oxidative phosphorylation and can survive damage to its mitochondrial DNA (mtDNA) ranging from simple mutations ('petite' phenotype) to complete loss of the mtDNA (ρ^0 phenotype) [3, 4]. The robustness of *S. cerevisiae* to mitochondrial mutations and harsh transformation protocols makes it one of the few organisms in which gene deletion or integration in the mtDNA is possible [5, 6]. Currently, most of these protocols rely on microprojectile bombardment using microscopic DNA-coated metal particles (biolistic transformation) to deliver DNA to the mitochondrial matrix [6, 7].
Despite intensive efforts, genetic engineering of *S. cerevisiae* mtDNA using RNA-guided endonucleases (RGENs) such as Cas9, has not yet demonstrated gene integration or deletion in mtDNA [8-11]. The inability to confidently edit mtDNA using RGEN systems is likely due to their three-component nature, as besides targeting the endonuclease to the mitochondria, it also requires targeting of two nucleic acids. A gRNA guides the endonuclease to the editing site and a DNA fragment repairs the dsDNA cut caused by the endonuclease for integration or deletion through homologous recombination. Targeting the endonuclease is straightforward, as 99 % of the mitochondrial proteome is synthesized in the cytosol and imported in the mitochondria and several robust mitochondrial targeting peptides are well characterized [12, 13]. However, despite reports of RNA targeting to the mitochondria [14-16], gRNA import in yeast mitochondria is still unreliable [17] and specific DNA import in eukaryotic mitochondria is likely complicated by the strong mitochondrial membrane potential and has not been demonstrated *in vivo* [18-20].

Efficient engineering of the mtDNA, via biolistic transformation or RGEN engineering, requires a method to select for transformants that carry the desired mutation in the mtDNA. To date, three approaches have been implemented for the selection of positive transformants. One strategy consists in repairing a ρ^- mutant carrying a deletion in one of the mitochondrial genes, typically a gene involved in oxidative phosphorylation. Positive mutants can then be identified by their ability to respire [6, 21]. The downside of this approach is that it requires the construction of a ρ^- strain with a defined mutation, and only few yeast strains are available with such mutations. The Barstar dominant toxin/antitoxin marker presents an alternative strategy that does not

require prior modification of the mtDNA. In this system an aspecific RNase is targeted to the mitochondria, and is counteracted by the successful expression of Barstar from the mtDNA [22].
Restoration of prototrophy in *arg8* deletion mutants auxotrophic for arginine is yet another elegant strategy to screen for successful editing of mtDNA. Acetylornithine aminotransferase, encoded by *ARG8*, catalyses the essential step from *N*-acetyl-L-glutamate-5-semialdehyde to *N*-acetyl ornithine in the mitochondrially localized arginine biosynthetic pathway (Figure 4.1) [23]. Deletion of *ARG8* from its native nuclear locus causes arginine auxotrophy, which can be complemented by integrating *ARG8* in the mtDNA. Mitochondrial expression of nuclear genes requires recoding to match the mitochondrial genetic code in which the UGA codon encodes a tryptophan, whereas UGA encodes a STOP codon in the nucleus [24]. This can be used as an additional layer of selection for correct mtDNA editing as genes that are codon-optimized for expression from the mtDNA will result in truncated proteins if expressed in the cytosol. [25]. The latter strategy has been widely used to study mitochondrial gene expression and splicing [25-34].

In an attempt to develop more accessible methods to engineer mitochondrial DNA in *S. cerevisiae*, we stumbled upon an unexpected and undesired observation: growth of *Δarg8* deletion mutants led to revertant mutants that were able to grow without arginine supply, making one of the very few selectable markers available for mtDNA-editing unusable. The present study identified the molecular basis underlying the recovery of arginine prototrophy in *Δarg8* mutants and used this information to construct robust arginine auxotrophs that can be reliably used for mitochondrial DNA engineering. Alternative auxotrophic markers that are suitable for recoded expression in the mitochondria were also explored.

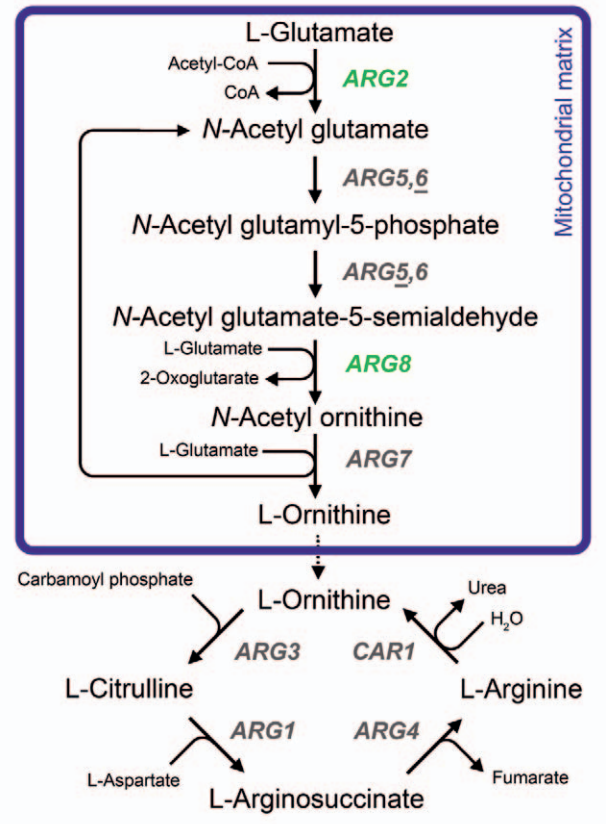


Figure 4.1. ***De novo*** **L-arginine biosynthesis in** ***S. cerevisiae*****.** *ARG2* (amino-acid *N*-acetyltransferase EC 2.3.1.1) and *ARG8* (acetylornithine aminotransferase EC 2.6.1.11) are subject of this study and indicated in green.

4.2 Materials & Methods

Strains & culture conditions

Saccharomyces cerevisiae strains constructed in this study were derived from the CEN.PK lineage [35]. Strains derived from the S288C lineage with a BY4741 background [36] were purchased from the Yeast Knockout (YKO) collection [37] (Horizon Discovery, Waterbeach, United Kingdom). Yeast strains (Table 4.1) were grown aerobically at 30 °C in 500 mL shake flasks containing 100 mL synthetic medium with ammonium as nitrogen source (SM) supplied with vitamins and trace elements, prepared and sterilized as described previously [38], in an Innova incubator shaker (Eppendorf, Hamburg, Germany) set at 200 rpm. Sterilized media were supplemented with a D-glucose solution to a final concentration of 20 g L^{-1} (SMD) or with absolute ethanol to a final concentration of 2 % (v/v) (SME). Glucose solutions (50 % w/v) were autoclaved separately for 10 minutes at 110 °C. Arginine excess concentrations and biomass yields on arginine were calculated according to Pronk [39] based on data by Oura [40], and low arginine media were supplied with a separately sterilized solution of L-arginine·HCl to a final concentration of 50 mg L^{-1} L-arginine, while high arginine media were supplied with 450 mg L^{-1} of L-arginine. Strains with other auxotrophic requirements were routinely grown on synthetic medium supplemented with separately sterilized solutions of uracil (Ura) to a final concentration of 150 mg L^{-1}, histidine (His) to a final concentration of 125 mg L^{-1}, leucine (Leu) to a final concentration of 500 mg L^{-1} and/or methionine (Met) to a final concentration of 100 mg L^{-1} [39] For cloning, YPD medium, containing 10 g L^{-1} Bacto Yeast extract, 20 g L^{-1} Bacto Peptone and 20 g L^{-1} glucose was used, and when necessary supplemented with 200 mg L^{-1} G418 (geneticin) or 200 mg L^{-1} hygromycin. For

Table 4.1. *S. cerevisiae* strains used in this study

Strain name	*Relevant genotype*	*Parental strain*	*Source*
CEN.PK113-7D	*MATa*	-	[35]
CEN.PK113-5D	*MATa ura3-52*	-	[35]
IMX581	*MATa can1::Cas9-natNT2 Δura3*	CEN.PK113-7D	[41]
IMX1714	*MATa sga1::Cas9-natNT2*	CEN.PK113-7D	[42]
IMX2100	*MATa sga1::Cas9-natNT2 Δura3*	IMX1714	[43]
IMX2600	*MATa can1::Cas9-natNT2*	CEN.PK113-7D	[44]
IMK703	*Δcar2*	IMX581	This study
IMK997	*Δarg2*	IMX1714	This study
IMK999	*Δarg8*	IMX1714	This study
IMK1031	*Δarg8 Δcar2 Δura3*	IMK1000	This study
IMC250	*Δarg8 Δcar2 Δura3 + pUDC434(preCOX4-ARG8 URA3)*	IMK1031	This study
BY4741	*MATa his3Δ1 leu2Δ0 met15Δ0 ura3Δ0*	-	[36]
BY4741 *KO1352 Δarg2*	*MATa his3Δ1 leu2Δ0 met15Δ0 ura3Δ0 arg2::KanMX*	BY4741	[37]
BY4741 *KO7711 Δarg8*	*MATa his3Δ1 leu2Δ0 met15Δ0 ura3Δ0 arg8::KanMX*	BY4741	[37]
IMS1186	Population #1 of IMK999 evolved for arginine prototrophy	IMK999	This study
IMS1187	Population #2 of IMK999 evolved for arginine prototrophy	IMK999	This study
IMS1188	Population #3 of IMK999 evolved for arginine prototrophy	IMK999	This study
IMS1190	Single-colony isolate of IMS1186	IMS1186	This study
IMS1191	Single-colony isolate of IMS1186	IMS1186	This study
IMS1192	Single-colony isolate of IMS1186	IMS1186	This study
IMS1193	Single-colony isolate of IMS1187	IMS1187	This study
IMS1194	Single-colony isolate of IMS1187	IMS1187	This study
IMS1195	Single-colony isolate of IMS1187	IMS1187	This study
IMS1196	Single-colony isolate of IMS1188	IMS1188	This study
IMS1197	Single-colony isolate of IMS1188	IMS1188	This study
IMS1198	Single-colony isolate of IMS1188	IMS1188	This study

plate cultivation 2 % (w/v) agar was added to the medium prior to heat sterilization. Single-colony isolates were obtained by restreaking single colonies three consecutive times on non-selective medium, except for IMC250 which carried a plasmid and was therefore restreaked on selective medium.
Frozen stocks were prepared by addition of sterile glycerol (30 % v/v) to late exponential phase shake-flask cultures of *S. cerevisiae* and 1 mL aliquots were stored aseptically at –80 °C.

Strain construction using Cas9-mediated DNA cleavage

Genomic mutations in the DNA using Cas9-mediated DNA cleavage and repair were performed as described previously by Mans *et al.* [45]. *URA3* deletion in IMX1714 and IMK1000 was performed using gRNA plasmid pUDR107 and annealed primers 13807 and 13808 as DNA repair (Table S4.1). Plasmids carrying gRNAs targeting *ARG2* and *ARG8* were transformed as linearized pMEL13 backbones and linear gRNA fragments with 60 bp overlapping ends to the linearized pMEL13 backbone, which assembled *in vivo* in yeast and simultaneously used for Cas9-mediated editing of the respective strains as described by Mans *et al.* [45]. For *CAR2* deletion, two gRNAs were first cloned into a pROS10 backbone using Gibson isothermal assembly as described ([45], Table S4.2). All gRNA and repair sequences were made by annealing 120 bp primers (Sigma Aldrich) and sequences are described in Table S4.1. Positive transformants were selected on appropriate selective media. Colonies were picked, resuspended in 15 µL of 0.02 M NaOH and boiled for 15 minutes at 99 °C to lyse cells for release of genomic DNA. 2 µL of the boiled cell suspension was used for diagnostic colony PCR using DreamTaq PCR Master Mix (Thermo Fisher Scientific, Landsmeer, The Netherlands) according to the manufacturer's instructions with primers described in Table S4.1. Positive transformants (Figure S4.1) were inoculated in 100 mL YPD and grown over 36 hours, then restreaked for single colonies on YPD solid medium to induce loss of the gRNA-plasmid. Colonies on YPD were screened for loss of their gRNA plasmid by restreaking on selective and non-selective YPD. Colonies that had lost the gRNA plasmid were restreaked again to obtain single-colony isolates, and frozen stocks were prepared as described earlier.

Plasmid construction & molecular biology techniques

Plasmid pUDC434 was cloned using Golden-Gate assembly performed according to the Yeast Toolkit (YTK) design of Lee *et al.* [46], from part plasmids pUD538, pYTK009, pGGkp321, pYTK053 (Table S4.2) and an *ARG8* gene (YTK part 3b), omitting the mitochondrial targeting sequence (MTS) as a linear DNA fragment. The *ARG8-NoMTS* was amplified from genomic DNA of CEN.PK113-7D isolated with the Qiagen Blood & Cell Culture DNA kit (Qiagen, Germantown, MD), following manufacturer's specifications, with Phusion High Fidelity Polymerase (Thermo Scientific) using primers (19477+19478) that were designed so the 20-aa MTS was omitted and the fragment was flanked with YTK-compatible Part 3b BsaI sites (Table S4.1). Part plasmid pGGKp321, containing the *preCOX4*-MTS [47] as part 3a, was PCR-amplified with Phusion polymerase using primers (17711 + 17712) from template pUDC286 so that the fragment was flanked with BsmBI and BsaI sites for constructing a Part 3a compatible with the YTK, and subsequently cloned into the YTK001-vector by Golden-Gate cloning, as described by Lee *et al.* [46]. All other part plasmids were obtained from the Yeast Toolkit collection [46]. 1 µL of the Golden-Gate reaction mixture was transformed into *E. coli* cells (XL1-Blue, Agilent Technologies, Santa Clara, CA), which were grown in solid Lysogeny broth (LB) medium (5.0 g L^{-1} yeast extract, 10 g L^{-1} Bacto tryptone, 2 % Bacto agar (BD Biosciences), 5.0 g L^{-1} NaCl) supplemented with 100 mg L^{-1} ampicillin or chloramphenicol. Upon growth, random colonies were picked and resuspended in 10 µL sterile water, of which 1 µL was used for diagnostic colony PCR, using DreamTaq MasterMix 2X (Thermo Fisher Scientific) and desalted primers (Sigma-Aldrich, Table S4.1), with an initial 10-minute incubation at 95 °C to release *E. coli* DNA. Positive transformants were inoculated in 5 mL liquid LB medium (omitting the Bacto agar) and plasmids were isolated using a GeneJET Plasmid Miniprep Kit (Thermo Fisher Scientific). Correct plasmid structure of isolated plasmids was confirmed by PCR using DreamTaq polymerase. Fragment sizes were assessed on a 1 % agarose gel.
100 ng pUDC434 (Table S4.2) was transformed into IMK1031 as previously described [48], mutants were selected on SMD medium. Mutants were restreaked three times on selective solid medium to obtain single colony isolates, which were grown and stored as described.

Growth rate analysis in microtiter plates

Growth rate analysis was performed in 96-wells microtiter plates at 30 °C and 250 rpm using a Growth Profiler 960 (EnzyScreen BV, Heemstede, The Netherlands). Frozen glycerol stocks were inoculated in 100 mL SMD + excess arginine, supplemented with Ura/His/Leu/Met if required, and grown overnight. 0.5 mL of the overnight culture was transferred to 100 mL SMD + excess Arg and grown until the OD_{660} had doubled at least once to ensure exponential growth. OD_{660} was measured using a Jenway 7200 spectrophotometer (Jenway, Staffordshire, UK) at 660 nm. The cultures were spun down and washed twice in sterile dH_2O, then diluted to an OD_{660} of 15. From this culture a 96-wells microtiter plate (EnzyScreen, type CR1496dl) containing SMD or SME either with excess or no arginine with final working volumes of 250 µL was inoculated with a starting OD_{660} of 0.3. Growth rate analysis and data analysis was performed as described by Boonekamp *et al.* [49]. For strains that did not reach an OD_{660} of 10, the growth rate was determined from the growth phase where the OD_{660} was higher than 1 and of which the slope of the natural logarithm of the OD_{660} measurements had an R^2 of at least 0.99, indicating exponential growth.

Growth analysis in shake flasks

Isolates from microtiter cultivation were analyzed in 500 mL shake flasks containing 100 mL SMD at 30 °C on an orbital shaker set at 200 rpm. A pre-culture was done on SMD with low arginine concentration (50 mg L^{-1}), and OD_{660} was measured at regular intervals. Once reaching stationary phase, cultures were washed with sterile dH_2O and transferred to medium without arginine or with excess arginine. To ensure growth on medium without arginine was not caused by arginine carryover from the preculture, cultures grown on medium without arginine were washed and transferred again upon reaching stationary phase. Metabolite concentrations were determined extracellularly, by centrifugation of 1 mL of culture for 3 min at 13.000 $\times$ *g* and analysis of the supernatant using high-performance liquid chromatography (HPLC) using an Aminex HPX-87H ion-exchange column operated at 60 °C with 5 mM H_2SO_4 as the mobile phase with a flow rate of 0.6 mL min^{-1} (Agilent, Santa Clara).

The linear relation between OD_{660} and cell number and OD_{660} and dry weight was experimentally obtained from measurements of aerobically grown CEN.PK113-7D on SMD. Cells were counted on the Coulter Z2 Analyser in five replicates of serial dilutions of OD_{660} values and dry weight was measured of 33 samples with different OD_{660} values as described earlier [38].

Cell count (cells mL^{-1}) = $6.560 \cdot 10^6 \times OD_{660} - 3.542 \cdot 10^5$ (R^2 of 0.999)

Dry weight (g L^{-1}) = $0.1353 \times OD_{660} + 0.0451$ (R^2 of 0.988)

DAPI staining and imaging

Nucleic acid staining and imaging was performed as described in Bouwknegt *et al.* [50]. In short, cultures were grown overnight on SMD + excess arginine media. Approximately 10^7 cells were harvested and washed with dH_2O, then resuspended in 1 mL of SMD and supplemented with 300 nM 4′,6-diamidino-2-phenylindole dihydrochloride (DAPI, Sigma Aldrich). The cells were incubated in the dark for 10–15 min at room temperature. Phase-contrast microscopy was performed using a Zeiss Axio Imager Z1 (Carl Zeiss AG, Oberkochen, Germany) equipped with a HAL 100 Halogen illuminator, HBO 100 illuminating system and AxioCam HRm Rev3 detector (60 N-C 1'' 1.0×) (Carl Zeiss AG). The lateral magnification oil objective 100×/1.3 was used with Immersol 518 type F immersion oil (Carls Zeiss AG). Fluorescence of DAPI dihydrochloride fluorescence, filter set 49 (excitation SP 380, emission BP 445/50) was used (Carl Zeiss AG). Results were analysed using the Fiji package of ImageJ [51]. Average cell fluorescence was measured as follows: using the phase-contrast image, an automatic threshold was set using the Otsu algorithm to detect cell outlines, which were stored as region of interest (ROI). Next, the background of the DAPI-image was removed with a rolling-circle radius of 20 pixels. The generated ROIs of each cell were overlaid on the DAPI image and the mean fluorescence of each ROI (cell) was measured.

Illumina Whole Genome Sequencing

Genomic DNA of CEN.PK113-7D, IMS1191, IMS1194 and IMS1197 was extracted with the Qiagen Blood & Cell Culture DNA kit (Qiagen, Germantown, MD), following manufacturer's specifications. Whole genome Sequencing (WGS) was performed using short-read sequencing and sequenced in-house on a MiSeq sequencer (Illumina, San Diego, CA) to obtain 151 cycle paired-end libraries with an insert-size of 550 bp through TruSeq DNA PCR-Free LT Library preparation (Illumina) using MiSeq Reagent Kit v3, yielding approximately 700 Mb of data per sample. All Illumina sequencing data are available at NCBI (https://www.ncbi.nlm.nih.gov/) under the Bioproject accession number PRJNA906461. Reads were mapped using BWA (version 0.7.15) [52] to a IMX2600 reference [44] Alignments were processed using SAMtools (version 1.3.1) [53], and variants were called by applying Pilon (version 1.18) [54]. The average coverage over the mitochondrial genome was determined by calculating the average amount of reads in the BWA-alignment in sliding windows of 500 base pairs. Sequencing output was visualized using the Integrated Genomics Viewer [55] and chromosome copy numbers were determined using Magnolya [56]. Multiple sequence alignments of protein and nucleotide sequences of *UME6* were performed using Clustal Omega (version 1.2.4) [57].

4.3 Results & Discussion

A suppressor mutation of an *ARG8* deletion rescues arginine prototrophy

The stability of Δ*arg8* mutants in the absence of arginine was assessed in microtiter plates. Using CRISPR-Cas9 editing, an Δ*arg8* mutant was constructed in the CEN.PK strain background (Figure S4.1). The knockouts were performed in a fully prototrophic strain IMX1714 (CEN.PK113-7D, expressing SpyCas9 [42]), leading to IMK999 (Δ*arg8*). The growth rate of the strain was determined in the presence and absence of arginine, in triplicate. In the presence of arginine, the Δ*arg8* strain grew as fast as its respective control strain CEN.PK113-7D (Figure 4.2A,B). In the absence of arginine, the Δ*arg8* mutant initially did not grow. However, after 290 hours, growth was observed in one of the three wells containing IMK999 and within 100 h, also the other two wells showed biomass growth with an average rate of 0.03 h^{-1} (Figure 4.2C).

This unexpected ability of Δ*arg8* mutants to grow in the absence of arginine could result either from adaptation to the arginine-free medium or from the occurrence of suppressor mutations. To test which mechanisms was involved, the Δ*arg8* mutant cultures, selected after growth on medium devoid on arginine shown in Figure 4.2, were transferred to shake flasks containing defined glucose medium (SMD) supplied with a low concentration of arginine (50 mg L^{-1} vs 450 mg L^{-1} in standard SMD). This medium composition was defined to ensure that arginine was the first depleted nutrient while all other nutrients were in excess, and therefore to minimize arginine carry-over during transfer to arginine-free medium. As *S. cerevisiae* biomass yield is 34 mg biomass per mg of arginine [40], the expected biomass concentration upon arginine depletion was approximately 1.7 g L^{-1} yeast dry weight, translating to an OD_{660} of 12. While typical arginine-rich cultures lead to max OD_{660} of 25 - 30, the Δ*arg8* mutants cultures reached a maximal OD_{660} of 10 - 12, indicating successful arginine depletion (Figure 4.2D). These arginine-depleted cultures were then transferred to shake flasks containing either no arginine or arginine excess (450 mg L^{-1}), in independent triplicate cultures (Figure 4.2D).

The cultures fed with excess arginine displayed a typical growth response, with specific growth rates around 0.4 h^{-1}, a final OD_{660} around 27 and complete utilization of glucose. Also, the ethanol produced during the glucose consumption phase was completely consumed (Figure 4.2E). The physiology of the Δ*arg8* mutants of IMK999 selected after growth on medium devoid on arginine, was different in the absence of arginine. The growth rate was around four-fold lower than arginine-rich cultures, the final OD_{660} was approximately five-fold lower, and while glucose was fully consumed, the produced ethanol remained untouched in the culture. To ensure that growth was not caused by arginine carry-over from the pre-cultures, the strains grown without arginine were transferred a second time in fresh SMD medium without arginine. This second transfer showed the same growth characteristics as the first transfer, confirming that the strains were indeed prototrophic for

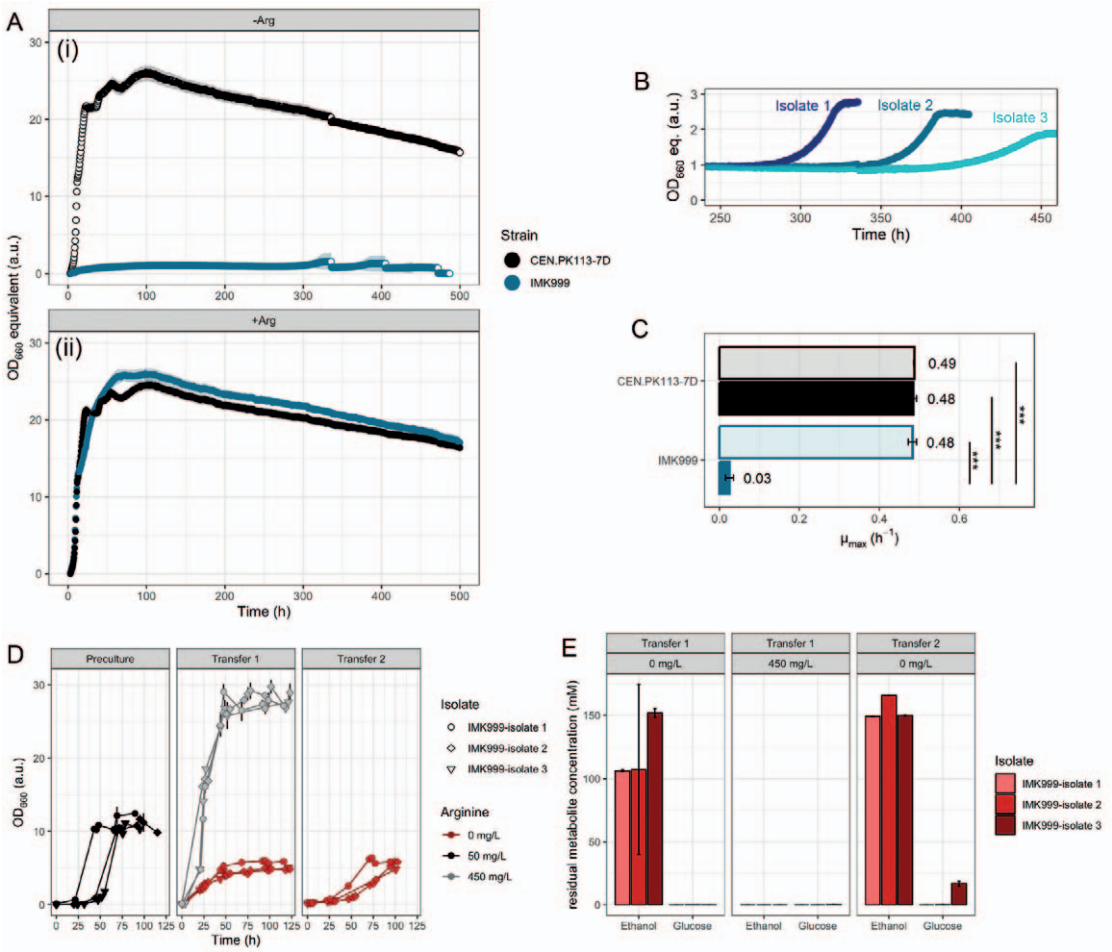


Figure 4.2. Growth of *Δarg8* mutants in the presence or absence of arginine. A) Growth of IMK999 and the prototrophic control CEN.PK113-7D in microtiter plates in the absence (-Arg, i) and presence (+Arg, ii) of arginine. Data represent the average and standard deviation of 3 independent culture replicates. B) Growth of three replicates of IMK999 in the absence of arginine in microtiter plates after 290 hours of growth C) Maximum specific growth rates (μ_{max}) calculated from the data shown in panel A. Data represent the average and standard deviation from triplicate growth experiments in presence (open bars) and absence (closed bars) of arginine. The growth rate of IMK999 in absence of arginine was measured after 300 h of lag phase. *** indicates a significant difference with a $p < 0.0005$ (two- tailed student T-test). D) Growth curves of three strains isolated from IMK999 (Δ*arg8*) cultures able to grow in the absence of arginine. The three isolates 1-3 were first grown in low arginine medium (50 mg L^{-1}, 'Preculture'), then transferred to high arginine (450 mg L^{-1}, gray, Transfer 1) and no arginine medium (0 mg L^{-1}, red, Transfer 1). The strains grown in absence of arginine were transferred a second time (Transfer 2) in arginine-free medium. E) Concentrations of glucose and ethanol in mM in the supernatant of the three isolates of IMK999 (in red) upon reaching stationary phase. The measurements were performed on samples taken at the final time points of the cultures shown in panel D.

arginine (Figure 4.2D). The most remarkable result was the ability of all three Δ*arg8* cultures to grow in the absence of arginine without lag phase, despite an intermediate transfer from medium supplied with arginine. This absence of lag phase strongly suggested the occurrence of suppressor mutations that conferred to Δ*arg8* mutants the constitutive ability to synthetize arginine.

To investigate whether recovery of arginine prototrophy was strain-specific, an *Δarg8* mutant with a BY4741 background from the widely used S288C lineage was also grown in the absence of arginine for a prolonged period of time. In this strain background, growth at a rate of 0.01 h^{-1} was detected after 200 hours (Figure S4.2). Recovery of arginine prototrophy in *Δarg8* mutants therefore seemed to be spread over different *S. cerevisiae* lineages. While the time frame required for the occurrence of suppressor mutations in microtiter plates might not seem relevant for standard strain construction programs, its frequency was rather high. Growth without

arginine of the Δ*arg8* mutants occurred in all tested cultures and strains after approximately 200 to 300 hours in microtiter plates (Figure 4.2C, Figure S4.2). The wells were inoculated with approximately 300 cells and since a stable Δ*arg8* mutant is not able to divide and therefore accumulate mutations, the mutations were most likely already present in the pre-culture used to inoculate the wells. Assuming a growth rate of approximately 0.03 h^{-1} for suppressor mutants (Figure 4.2B), it would take a single cell approximately 300 hours to reach an OD_{660} of 1. Extrapolating this numbers to the observed growth in microtiter plates, it means that at least one suppressor mutant was inoculated in each well containing 300 cells, leading to a mutation frequency of at least 1 in 300. Working with larger culture volumes than the 250 µL microtiter plates, as typically done in strain construction programs, will necessarily increase the number of suppressor mutants at the start of the culture, and thereby the occurrence of arginine prototrophic Δ*arg8* cultures. This high frequency of occurrence of suppressor mutants therefore underlines the importance of identifying a mitochondrial selection marker more stable than *ARG8*.

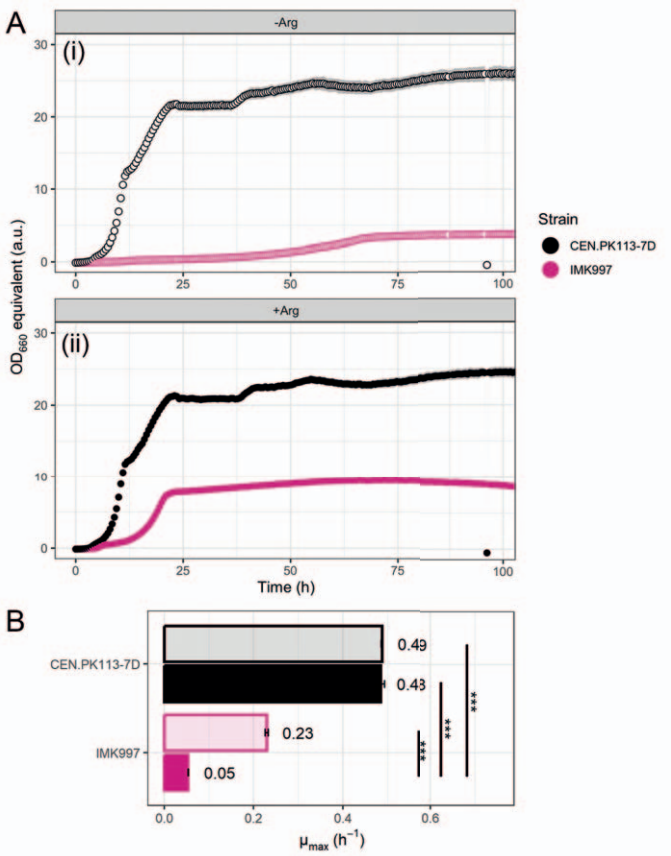


Figure 4.3. Growth of *Δarg2* mutants in the presence and absence of arginine. A) Growth of IMK997 and the prototrophic control CEN.PK113-7D in the absence (-Arg, i) and presence (+Arg, ii) of arginine in microtiter plates. Data represent the average and standard deviation of three independent culture replicates. B) Maximum specific growth rates (μ_{max}) of IMK997 and CEN.PK113-7D in the presence (open bars) and absence (closed bars) of arginine. The data represent the average and standard deviation of independent culture replicates. *** indicates a significant difference with a $p < 0.0005$ (two- tailed student T-test).

Search for alternative mitochondrial auxotrophic markers

As Δ*arg8* mutants were unstable, other potential auxotrophic mitochondrial markers were investigated. In addition, mitochondrial gene engineering would benefit from the availability of multiple auxotrophic markers. Enzymes involved in the biosynthesis of branched chain amino acids, lysine, and arginine, which have (a part of) their biosynthetic routes localized to mitochondria [23, 58, 59] were screened at the *Saccharomyces* Genome Database (SGD: https://www.yeastgenome.org [60]) for their suitability as mitochondrial auxotrophic marker (Table S4.3). The requirements for these auxotrophic markers were i) the lack of known suppressor mutations, ii) the availability of permissive conditions in which the auxotrophic mutant can grow at (near) wild-type rates, iii) the specific localization in mitochondria and no other compartment. Six enzymes met all the requirements, *ILV2*,*3*,*5*, *BAT1*, *ARG2* and *LYS12* (Table S4.3). *ILV2*,*3*,*5* and *BAT1* encode enzymes involved in branched-chain amino acid biosynthesis in which they catalyse multiple parallel reactions. Deletion of any of these enzymes results in a double auxotrophy for isoleucine and valine [59], which complicates screening as multiple pathways are impacted. Little information is available on the phenotype of *LYS12* null mutants, but the absence of any tryptophan (Trp) residues in *LYS12* makes it less attractive as mitochondrial selection marker. Deletion of *ARG2* was described to result in arginine (near-) auxotrophy [61, 62]. *ARG2* contains five Trp (in comparison to the two Trp in *ARG8*), that can be replaced by the UGA codon that encodes a Trp in the

mitochondria and a stop codon in the cytosol, making *ARG2* suitable for use as a mitochondrially-targeted auxotrophic marker.

Using CRISPR-Cas9 editing, an Δ*arg2* mutant was constructed in the CEN.PK strain background. The knockouts were performed in a prototrophic strain (IMX1714). The growth rates and growth profiles of the mutants were measured in microtiter plates in the presence and absence of arginine to check their arginine requirements, together with the prototrophic control strain CEN.PK113-7D.
In contrast to earlier reports, the *Δarg2* mutant (IMK997) was able to grow reasonably well in the complete absence of arginine, albeit with a 2 to 3-fold lower final OD_{660} and a ca. 5-fold lower specific growth rate than in the presence of arginine, and without extensive lag phase (Figure 4.3). The specific growth rate of 0.05 h^{-1} of the Δ*arg2* mutant revealed its ability to synthetize arginine at slow rates. To assess if this phenotype was specific to the CEN.PK lineage or conserved between yeast strains, an *Δarg2* mutant with an S288C background was also grown in the absence of arginine. This mutant grew similarly to the CEN.PK *Δarg2* mutant with a growth rate of 0.06 h^{-1} in the absence of arginine (Figure S4.2), indicating that the leaky arginine auxotrophy of *Δarg2* mutants was not strain-dependent. *ARG2* encodes *N*-acetylglutamate synthase, first step in the arginine biosynthetic pathway that acetylates L-glutamate to form N-acetyl glutamate using acetyl-CoA as acetyl donor (Figure 4.1). There is no other known enzyme dedicated to the acetylation of glutamate in *S. cerevisiae*. However, ornithine acetyltransferase (encoded by *ARG7*) that catalyses the last step in ornithine synthesis has been shown to

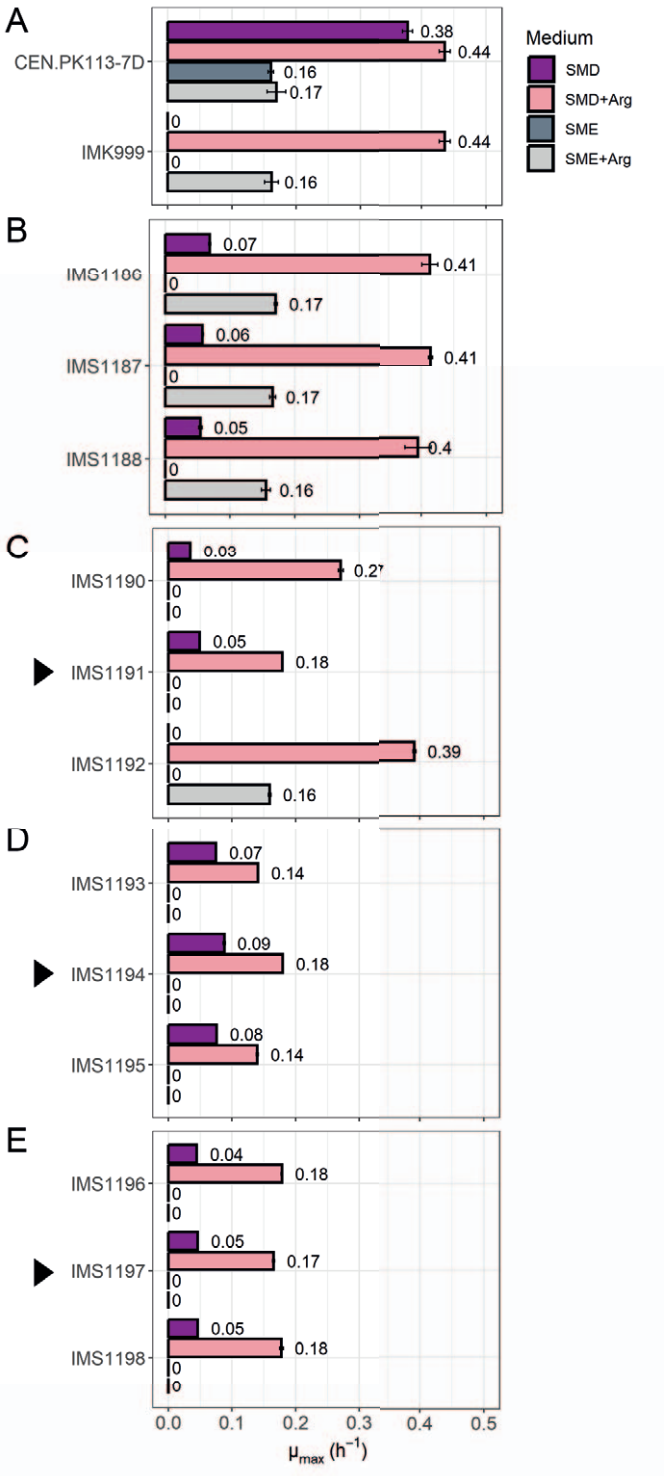


Figure 4.4. Specific growth rate of evolved strains with *Δarg8* background able to grow without arginine and their control. A) Prototrophic control CEN.PK113-7D and Δ*arg8* parent IMK999. B) Evolved arginine prototrophic populations IMS1186, IMS1187 and IMS1188. C) Single colony isolates IMS1190-1192 derived from IMS1186. D) single colony isolates IMS1193-1195 derived from IMS1187. E) Single colony isolates IMS1196-1198 derived from IMS1188. The specific growth rate was determined in microtiter plates on SM-glucose (SMD) or SM-ethanol (SME) in absence or presence of arginine (Arg). Data represent the average and standard deviation of independent culture triplicates. Arrows indicate isolates that were used for whole genome sequencing, full growth profiles are shown in figures S4.2- 4.4.

have residual acetylglutamate synthase activity. Overexpression of *ARG7* in an *Δarg2* mutant increases acetylation of L-glutamate and enables growth in the absence of arginine [63]. It is therefore likely that the deletion of *ARG2* in the CEN.PK background is complemented by the L-glutamate acetyltransferase activity of the *ARG7*-encoded ornithine acetyltransferase. In this scenario, the double deletion of *ARG7* and *ARG2* would therefore be required to fully abolish arginine synthesis. Regardless of the underlying mechanism, the ability of *Δarg2* mutants to grow in the absence of arginine disqualified *ARG2* as mitochondrial selection marker.

The gain of arginine prototrophy in *Δarg8* mutants comes at the cost of respiration

In the absence of viable alternative to *ARG8*, identifying and eliminating the mechanism leading to arginine synthesis in the *Δarg8* mutant would restore *ARG8* as a potential mitochondrial selection marker. The *Δarg8* mutants were able to grow in the absence of arginine, even after multiple transfers and an intermediate transfer on medium with arginine. This indicated the occurrence of mutations enabling arginine biosynthesis, rather than simple acclimation to the arginine-free medium (Figure 4.2E). As evolution is a stochastic process, with continuous occurrence, fixation and loss of random mutations, the three evolved populations prototrophic for arginine started from IMK999 (Figure 4.2C,D) most likely contained a mixture of 'true' prototrophs and 'cheaters' that are auxotrophic for arginine but cross-feed from the prototrophic mutants [64, 65]. Additionally, the population was also most likely heterogenous by the molecular mechanisms underlying arginine prototrophy. To assess heterogeneity regarding arginine requirements in the *Δarg8* population and identify true arginine prototrophs, the three *Δarg8* cultures prototrophic for arginine (henceforth indicated as evolved cultures IMS1186, IMS1187 and IMS1188) were restreaked on non-selective medium. Three single colony isolates were selected for each of the three cultures, leading to a collection of nine individual *Δarg8* strains named IMS1190 to IMS1198. The evolved populations (IMS1186, IMS1187 and IMS1188) and the nine single colony isolates were tested for their ability to grow without arginine on arginine-free SM with glucose as carbon source. The prototrophic strain CEN.PK113-7D and the *Δarg8* parental strain IMK999 were included as control strains. As expected, CEN.PK113-7D grew with and without arginine at 0.4 h^{-1} (Figure 4.4A, Figure S4.3). IMK999 grew as fast with arginine but was unable to grow without arginine in the running time of the experiment (140 h). The three evolved populations IMS1186, IMS1187 and IMS1188 also grew at the same growth rate as CEN.PK113-7D and IMK999 in the presence of arginine (Figure 4.4B, Figure S4.3), and grew at specific rates between 0.05 and 0.07 h^{-1} in the absence of arginine. As the arginine prototroph *Δarg8* mutants seemed to have lost their ability to consume ethanol (Figure 4.2E), growth was also tested on medium with ethanol as sole carbon source. The three evolved populations IMS1186, IMS1187 and IMS1188 were indeed unable to grow on ethanol as sole carbon source in the absence of arginine, but grew as fast the control strains CEN.PK113-7D and IMK999 (approximately 0.16 h^{-1}) in the presence of arginine.

The nine single colony isolates displayed growth profiles different from their parental populations, confirming that IMS1186, IMS1187 and IMS1188 were mixed populations of evolved and unevolved cells (Figure 4.4C,D,E, Figure S4.4, Figure S4.5). Two different phenotypes were observed among the nine isolates. IMS1192 grew as fast as the controls in the presence of arginine and was auxotrophic for arginine, suggesting that it was genetically identical to the *Δarg8* auxotrophic strain. This isolate was also the only strain able to utilize ethanol as carbon source in the presence of arginine. The eight other isolates grew slower than the IMS1186, IMS1187 and IMS1188 evolved populations on medium with glucose and in the presence of arginine, with growth rates of approximately 0.17 h^{-1}, with the exception of IMS1190, which grew faster on this medium (0.27 ± 0.004 h^{-1}, Figure 4.4C). In the absence of arginine, these isolates grew at specific growth rates between 0.03 and 0.09 h^{-1}. None of these strains were able to grow on ethanol, neither in the presence nor in the absence of arginine. These results suggested that the acquisition of arginine prototrophy coincided with a loss of the ability to consume ethanol. As ethanol consumption requires respiration, this response might hint at the loss of respiratory function in the arginine prototrophic *Δarg8* mutants.

To identify the molecular basis for arginine prototrophy and loss of ethanol utilization, the genome of three *Δarg8* prototrophic mutants IMS1191, IMS1194 and IMS1197, isolated from the three different evolved populations, were sequenced. While some genome rearrangements were found in the different strains (duplication of chromosome IX in IMS1191 and of chromosome V in IMS1197 (Table 4.2, Figure S4.6 - Figure S4.9), the only genomic alteration shared by the three strains was a decrease in mitochondrial DNA (mtDNA) coverage.

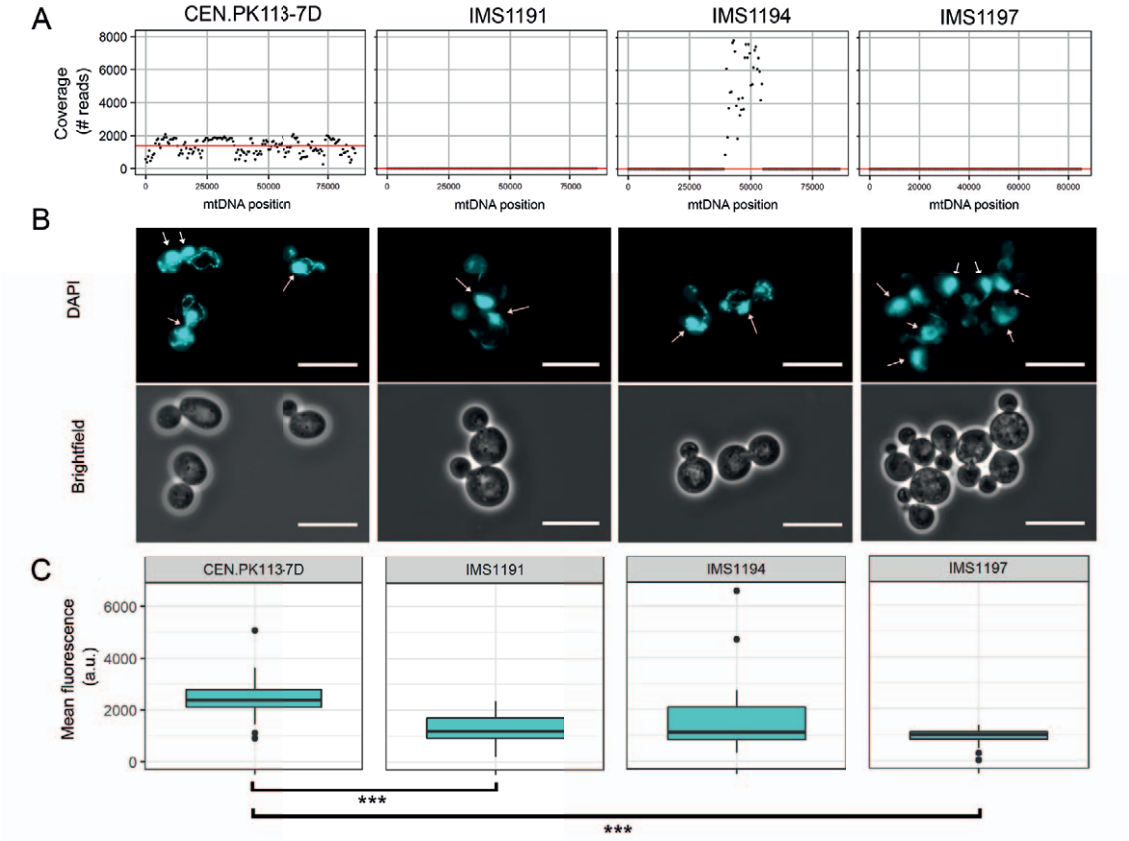


Figure 4.5. Characterization of the mitochondrial DNA content of CEN.PK1137D, IMS1191, IMS1194 and IMS1197. A) Sequencing coverage depth of the mitochondrial genome. B) Fluorescence and brightfield microscopy of nuclei and mitochondrial nucleoids of yeast cells stained with the DNA-specific fluorescent dye DAPI, at a 1000 × magnification. Scale bars represent 10 µm. Nuclei are indicated with arrows; other nucleic acids are localized in the mitochondria. C) Mean fluorescence of DAPI per cell as a measure of DNA content from fluorescence microscopy pictures. The median value is represented by a horizontal line while the boxes represent the 25 and 75 % quartiles. Outliers are shown as dots. *** indicates a significant difference with a $p < 0.0005$ (two- tailed student T-test).

mtDNA was not found in IMS1191 and 1197, and only partial coverage was recovered for IMS1194 mtDNA (Figure 4.5A). Out of the 86 kb mitochondrial genome only 15 kb was recovered in IMS1194, resulting in a loss of all mitochondrially encoded genes with the exception of *OLI1* and *VAR1* (Figure 4.5A). Cellular DNA content was also monitored by fluorescence microscopy using the DNA-binding dye DAPI. DAPI binds DNA both in the nucleus and in mitochondria, but mtDNA can be visualized and quantified. Microscopic analysis of DAPI staining confirmed the absence of mtDNA in IMS1191 and IMS1197, and the lower abundance of mtDNA in IMS1194 (Figure 4.5B, 5C). As the mtDNA encodes essential subunits of proteins involved in oxidative phosphorylation [66], its (partial) loss was most likely the cause for the loss of respiratory capacity, and thereby ethanol consumption, in the *Δarg8* arginine prototrophic strains.

Arginine prototrophy is the caused by loss of function of Ume6

As the loss of mtDNA and respiratory capacity could not explain arginine prototrophy, single nucleotide polymorphisms (SNPs) and insertion-deletions (INDELs) that would lead to non-synonymous mutations in coding regions of the three mutants were analysed. This pointed to *UME6* as a potential culprit, as it contained an insertion in all three evolved strains (Table 4.2).

In IMS1191, an insertion of two base pairs caused early termination of the 836 amino acids-long Ume6 protein at 265 amino acids. In IMS1194 and IMS1197 an insertion of two base pairs caused a frame shift at

Table 4.2. Non-synonymous or non-silent mutations and genomic rearrangements following from whole genome sequencing of isolates IMS1191, IMS1194 and IMS1197

Strain	Chr	Position (bp)	Gene name	Mutation type	Location CDS (bp)	Mutation (bp)	Mutation (amino acid)
IMS1191	*IV*	847,234	*UME6*	Insertion	650	AT -> AAT	Frame shift +1, termination at 265 aa
	VII	371,208	*GUP1*	SNP	968	TTC ->TAC	Phe-323-Tyr
	IX			Duplication			
	XII	1,098,556	*YRF1-4*	SNP	1233	ATG -> ATA	Met-411-Ile
	XII	1,098,682	*YRF1-4*	Insertion	1336	G -> GA	Frame shift +1, early termination
	XIII	209,656	*RPS17A*	SNP	38	TCT -> TGT	Ser-13-Cys
	XIII	416,582	*VBA1*	SNP	305	TCA ->TAA	Ser-102-STOP
	mtDNA		Full loss of mtDNA (ρ^0)				
IMS1194	*III*	11,448	*VBA5*	SNP	1324	TTC -> ATC	Phe-442-Ile
	IV	845,375	*UME6*	Insertion	2509	AT -> AAT	Frame shift +1, extended by 4 aa
	XI	187,373	*SDH3*	SNP	338	CTC -> CCC	Leu-113-Pro
	XV	274,288	*IFM1*	SNP	1463	GAA -> GGA	Glu-488-Gly
	XV	824,641	*VPH1*	SNP	1162	GCT -> CCT	Ala-388-Pro
	mtDNA		Partial loss of DNA (ρ^-)				
IMS1197	*IV*	845,375	*UME6*	Insertion	2509	AT -> AAT	Frame shift +1, extended by 4 aa
	V			Duplication			
	VII	1,059,537	*YTA7*	SNP	3776	ACG -> ATG	Thr-1259-Met
	XII	1,098,556	*YRF1-4*	SNP	1233	ATG -> ATA	Met-411-Ile
	mtDNA		Full loss of mtDNA (ρ^0)				

the C-terminus of the protein, resulting in the extension of the protein with four amino acids (Figure S4.10, Figure S4.11). Ume6 has a primary role in the transcriptional regulation of early meiotic genes by facilitating DNA-binding of the Ume6-Sin3-Rpd3 complex [67-70]. Additionally, it is involved in the transcriptional regulation of genes involved in arginine catabolism and DNA-repair, that like early meiosis-specific genes, contain *URS1* sequence elements that Ume6 binds to [71-73]. The truncation in IMS1191 resulted in loss of the Ume6 repression domain and therefore loss of Ume6 repression activity in this strain [70]. The extension of Ume6 by four amino acids in IMS1194 and IMS1197 has a less obvious consequence, but since the phenotype is shared between all three strains, this mutation has likely resulted in a loss of function of Ume6. The 77 C-terminal amino acids of Ume6 contain the DNA-binding domain of the protein [73-75], which could be disrupted by extension of the protein. Ume6 represses arginine catabolism by repression of arginase (*CAR1*) and ornithine transaminase (*CAR2*) when preferred nitrogen sources are abundant in the culture medium (Figure 4.6A) [71, 73, 76]. The role of ornithine transaminase has been mostly studied in the context of arginine catabolism catalysing the synthesis from L-glutamate-5-semialdehyde from ornithine [77]. However, it has been shown in human cells that the reverse reaction, leading to the synthesis of ornithine is possible [78]. Additionally, overexpression of *CAR2* can alleviate the arginine auxotrophy of *Δarg8 S. cerevisiae* mutants [34]. It is therefore very likely that inactivation of Ume6 in the evolved *Δarg8* strains alleviated the repression of its targets, and more specifically *CAR2*, leading to ornithine synthesis from L-glutamate-5-semialdehyde in the cytosol and thereby to arginine synthesis.

The inactivation of Ume6, which supposedly occurred in amino acid-rich environments, has likely led to growth in the absence of arginine. However, due to its role in many metabolic- and signalling pathways, inactivation of the protein might also have detrimental effects on cell growth. For example, inactivation of Ume6 also leads to derepression of arginase (*CAR1*), that catalyses the conversion of L-arginine to L-ornithine (Figure 4.6A). To prevent the recycling of arginine back from ornithine, arginase reversibly but tightly binds to L-ornithine carbamoyltransferase (encoded by *ARG3*), leading to ornithine carbamoyltransferase inhibition without affecting arginase activity [79]. However, the ornithine carbamoyltransferase activity is essential for arginine synthesis, so the inhibition by arginase must be released in order to maintain arginine prototrophy. The binding between Car1p and Arg3p is sensitive to ornithine and arginine concentrations in the cytosol. Therefore, it is conceivable that these metabolites are maintained at low concentration in the evolved strains, enabling cells to retain sufficient L-ornithine carbamoyltransferase activity for arginine synthesis. Additionally, this proposed remodelled pathway for arginine synthesis present in the evolved strains, entails simultaneous expression of arginase and its antagonistic reaction encoded by *ARG1* (argininosuccinate

synthetase, Figure 4.6A). The reaction catalysed by Arg1p is ATP-consuming, resulting in an ATP-wasting futile cycle when arginase is not repressed. Therefore, constitutive Ume6 inactivation, while conferring arginine prototrophy, is most likely also detrimental to the evolved strains.
The loss of mtDNA is probably another side-effect of Ume6 inactivation in the arginine prototrophic *Δarg8* mutants. The deletion of *UME6* leads to the loss of respiratory activity or decreased fitness in respiratory media [80-83], however the underlying molecular mechanism has not been elucidated yet. Ume6 is a pleiotropic regulator that targets approximately 220 genes and acts both as activator and repressor [69]. The only known connection between Ume6 and the mtDNA is through its role in mitophagy, the selective autophagy of mitochondria. In particular, Ume6 represses *ATG8* and its target *ATG32*, a key gene in the initiation of mitophagy [84, 85]. The deactivation of Ume6 might therefore alter the regulation of mitophagy and result in loss of mitochondrial DNA in IMS1191, IMS1194 and IMS1197.

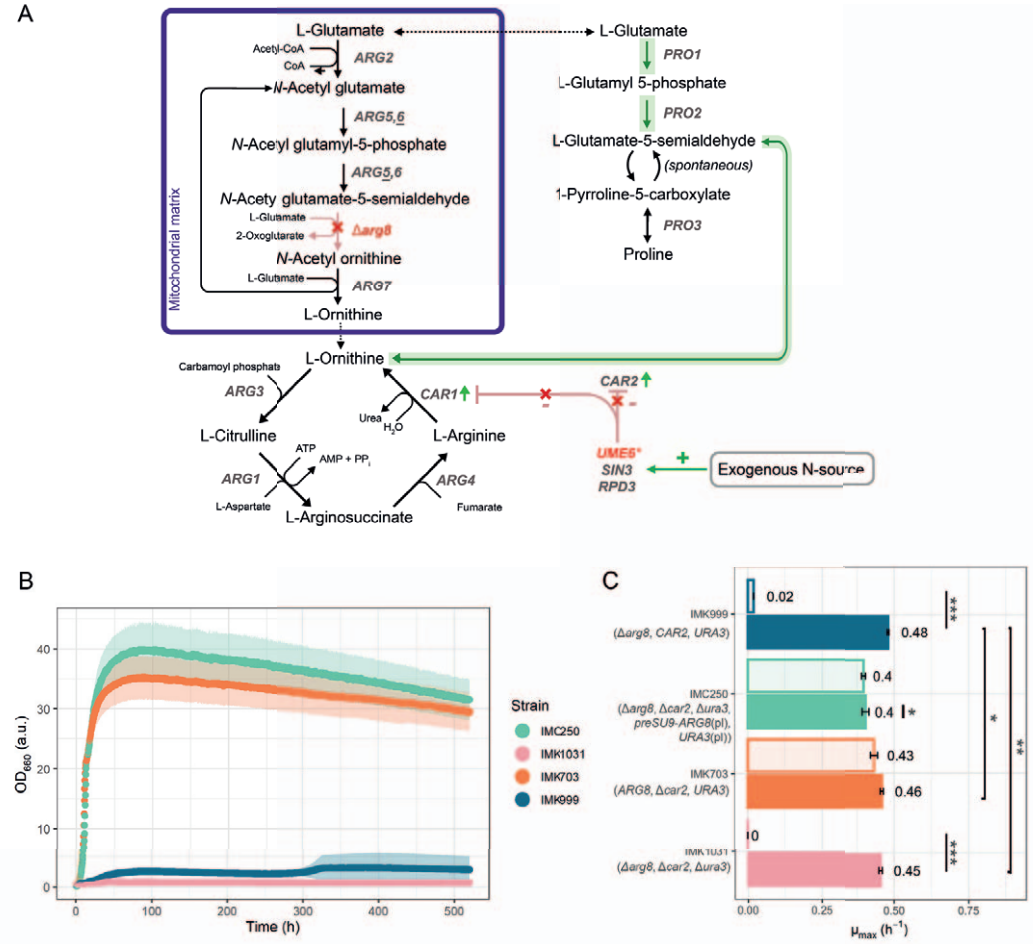


Figure 4.6. Characterization of *CAR2* deletion mutants. A) Proposed mechanism for arginine biosynthesis in evolved *Δarg8* deletion strains. Red lines indicate transcriptional repression by Ume6 on *CAR1* (arginase, EC 3.5.3.1) and *CAR2* (L-ornithine transaminase, EC 2.6.1.13) in the presence of exogenous nitrogen (N) sources. The proposed route for arginine synthesis in evolved *Δarg8* mutants is highlighted in green. B) Growth profiles of arginine-free cultures (in microtiter plates) of strains with single *CAR2* deletion (IMC250, *Δarg8 Δcar2 Δura3*, plasmid-based expression of preCOX4-*ARG8*, *URA3* and IMK703, *Δcar2*), single *ARG8* deletion (IMK999, *Δarg8*) and double *CAR2* and *ARG8* deletion (IMK1031, *Δarg8 Δcar2 Δura3*). The data represent the average and standard deviation of independent triplicate cultures; the standard deviation is indicated as a lighter colored shaded area. Growth profiles of all strains and controls are shown in Figure S4.12. C) Maximum specific growth rates (μ_{max}) in the presence (closed bars) and absence (open bars) of arginine. The growth rate of IMK999 was determined after 300 h of lag phase. Asterisks indicate a significant difference where a *** - $p < 0.0005$, ** - $p < 0.005$ and * - $p < 0.05$ (two- tailed student T-test). Significance was tested within strains on + and – arginine and between strains on + arginine. The growth rate of IMC250 is significantly the lowest of all strains on media with arginine.

Altogether, these side effects caused by the Ume6-dependent recovery of arginine prototrophy might contribute to the substantially slower specific growth rate (ten to thirteen-fold, Figure 4.4) and six- to seven-fold lower final OD (Figure S4.3 – 4.5) of the evolved *Δarg8* mutants in the presence of arginine, as compared to their arginine prototrophic parent.

Engineering of a stable arginine auxotrophic strain

Sequencing data suggested that derepression of *CAR2* might enable cytosolic arginine synthesis from L-glutamate-5-semialdehyde, an intermediate of proline synthesis. If this theory were true, double deletion of *ARG8* and *CAR2* would completely abolish arginine biosynthesis. As expected, the deletion of *CAR2* in a control, arginine prototrophic strain (strain IMK703) did not affect yeast physiology in the presence or absence of arginine (Figure 4.6B,C, Figure S4.12). The deletion of *CAR2* in the arginine auxotroph *Δarg8* (strain IMK1031) successfully prevented the appearance of arginine prototrophic mutants. Indeed, while growth without arginine was observed after ca. 300 hours of culture with *Δarg8* mutants (strain IMK999), no growth was observed after 500 hours with the *Δarg8Δcar2* double mutant (IMK1031, Figure 4.6B,C). Accordingly, plasmid-based expression of mitochondrially-targeted Arg8p in the *Δarg8Δcar2* double mutant restored arginine synthesis (strain IMC250, Figure 4.6B,C). The *Δarg8Δcar2* genetic background combined with the *ARG8* selection marker therefore offers a stable system for genetic engineering of yeast mitochondrial DNA.

4.4 Conclusions

Although *ARG8* is a widely used and popular auxotrophic marker for mitochondrial genome editing, physiology and protein import, the present work demonstrates the frequent occurrence of suppressor mutations (approximately 1 in 300 cells). The loss of function mutation in the *UME6* gene and the resulting transcriptional activation of *CAR2* are proposed as main molecular mechanisms for the recovery of arginine prototrophy in *Δarg8* mutants.

Remarkably, while the functional expression of *CAR2* should be sufficient to restore arginine synthesis in *Δarg8* mutants, arginine prototrophy in all three sequenced revertants was the result of a loss of function of Ume6. Considering the pleiotropic role of this master regulator [69], this evolutionary solution likely came with deleterious side effects, for instance by waste of resources expressing unnecessary proteins, deregulation of pathways such as mitophagy, loss of respiration, and activation of futile cycles. While multiple mutations might be required to enable expression of *CAR2* alone without affecting Ume6 (e.g., mutations in the *CAR2* promoter preventing Ume6 binding), a single mutation was sufficient to inactivate Ume6, thereby offering an easier evolutionary path for arginine synthesis despite detrimental side-effects. Recovery of prototrophy in the *Δarg8* mutants was observed during growth on glucose, where respiration is not required. It would be particularly interesting to explore the evolutionary path taken by Δ*arg8* mutants in ethanol media, in which loss of mtDNA and oxidative phosphorylation, and thereby Ume6 inactivation, is lethal.

The genes and pathways involved in the suppressor mutants are highly conserved among different strains of *S. cerevisiae*. The recovery of arginine prototrophy in Δ*arg8* mutants was indeed observed in both the CEN.PK and S288C lineages and is bound to happen in any *Saccharomyces cerevisiae* strain lineage. The suppressor mutation can be circumvented by additional deletion of *CAR2*, and the described *Δarg8Δcar2* double deletion provided stable arginine auxotrophy and prevented leaky arginine synthesis. When characterizing growth of strains with *Δarg8* background or performing laboratory evolution experiments, it is recommended to prevent false positives through deletion of *CAR2* or assessment of the integrity of *UME6*.

Acknowledgements

We thank Agostina Crotta Asis for assistance with DNA isolation and Illumina sequencing, Marcel van den Broek for processing raw sequencing data, Marijke Luttik for growth characterization of BY4741 strains, Sophie de Valk for measuring the OD/dry weight/cell count correlation and Angel Sevilla Camins for construction of strain IMK703 and plasmid pUDR131.

4.5 Supplementary data

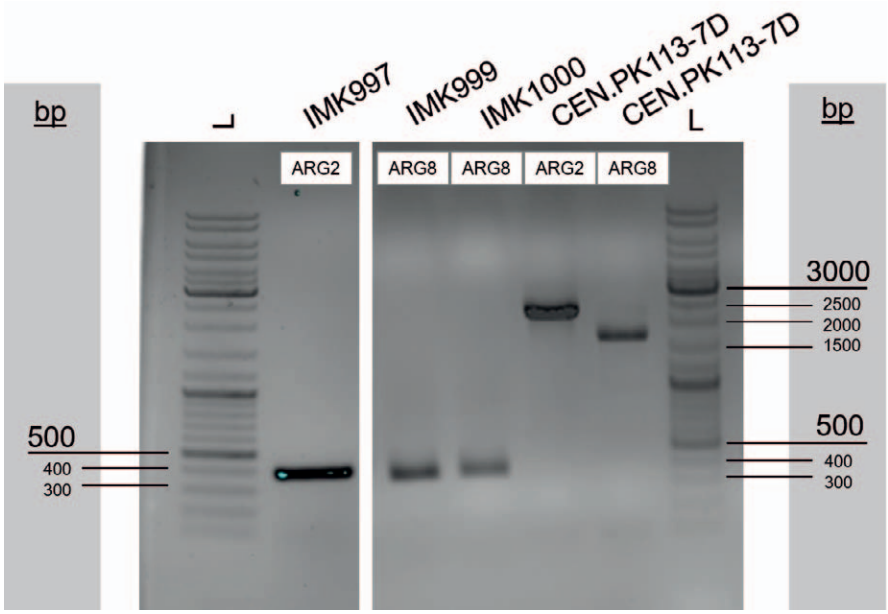


Figure S4.1. Confirmation of deletion of *ARG2* and *ARG8.* The respective loci were amplified using PCR and DNA products were separated on a 1 % agarose gel with a GeneRuler ladder Mix (L) as a reference. Relative basepair (bp) sizes are indicated at the left and right. Template DNA was isolated from the strains indicated (IMK997 - Δ*arg2*, IMK999,1000 - Δ*arg8*, CEN.PK113-7D – wildtype control). The white boxes indicate the amplified locus in each reaction, the same primer pairs were used for the same loci. Expected sizes: *ARG2* – 2111 bp, *ARG8* – 1584 bp, Δ*arg2* – 386 bp, Δ*arg8* – 312 bp.

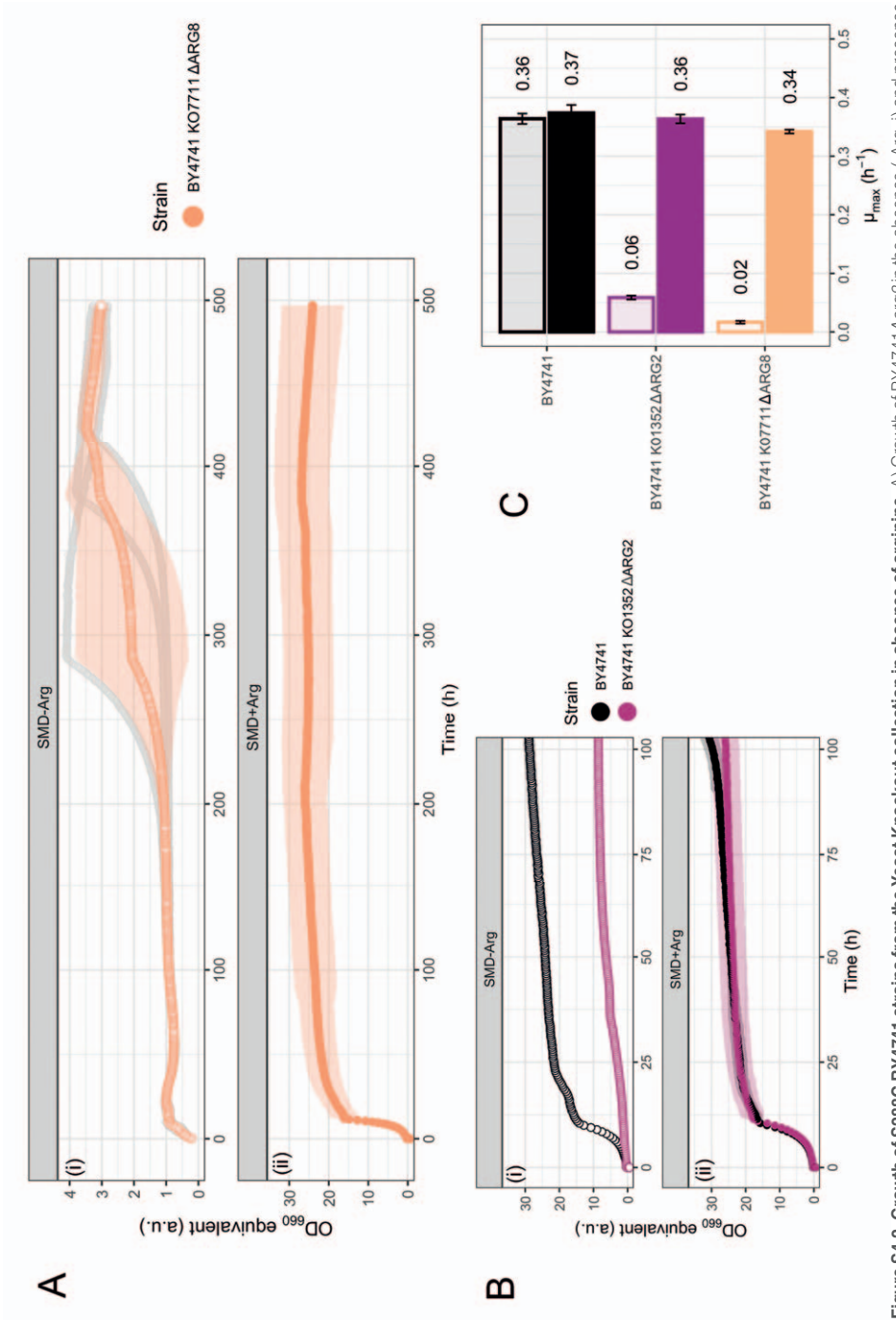


Figure S4.2. Growth of S288C BY4741 strains from the Yeast Knockout collection in absence of arginine. A) Growth of BY4741Δ*arg8* in the absence (-Arg, i) and presence (+Arg, ii) of arginine in microtiter plates. Orange data represent the average and standard deviation of 3 independent culture replicates, which are indicated in grey. B) Growth of BY4741Δ*arg2* and the prototrophic control BY4741 in the absence (-Arg, i) and presence (+Arg, ii) of arginine in microtiter plates. Data represent the average and standard deviation of 3 independent culture replicates. C) Maximum specific growth rates (μ_{max}) of BY4741Δ*arg2* BY4741Δ*arg8*, and BY4741 in the presence (open bars) and absence (closed bars) of arginine. The data represent the average and standard deviation of independent culture replicates.

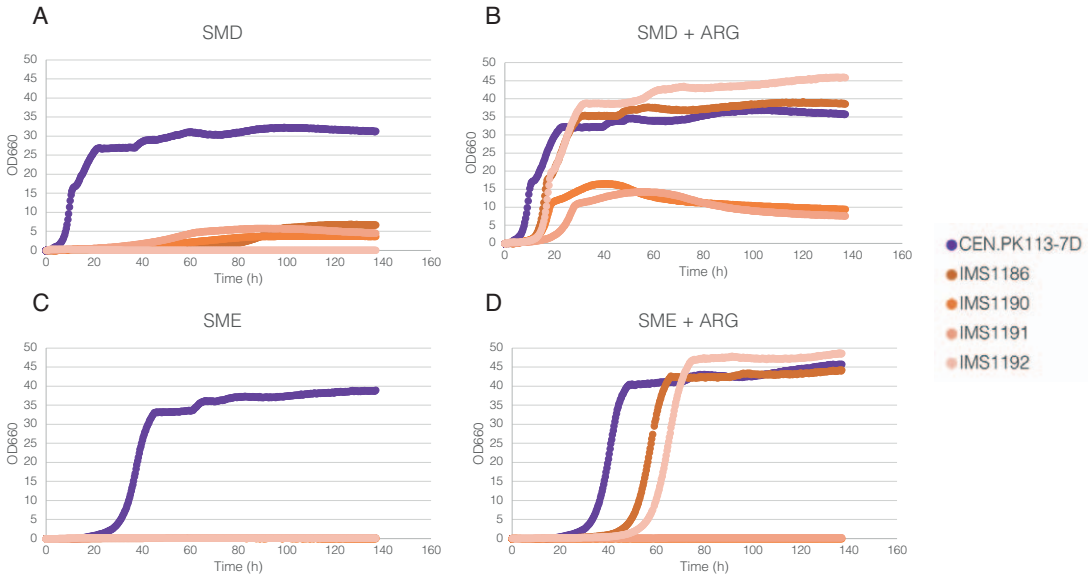


Figure S4.3. Growth of CEN.PK113-7D (prototrophic, purple), the evolved arginine prototrophic population IMS1186 (dark orange) and its three derived single-colony isolated IMS1190, 1191 and 1192 (light orange) in different media. Growth was characterized in microtiter plates on SM-glucose (SMD, A), SMD + arginine (B), SM-ethanol (SME, C) and SME + arginine (D). The data represent the average and standard deviation of independent culture replicates.

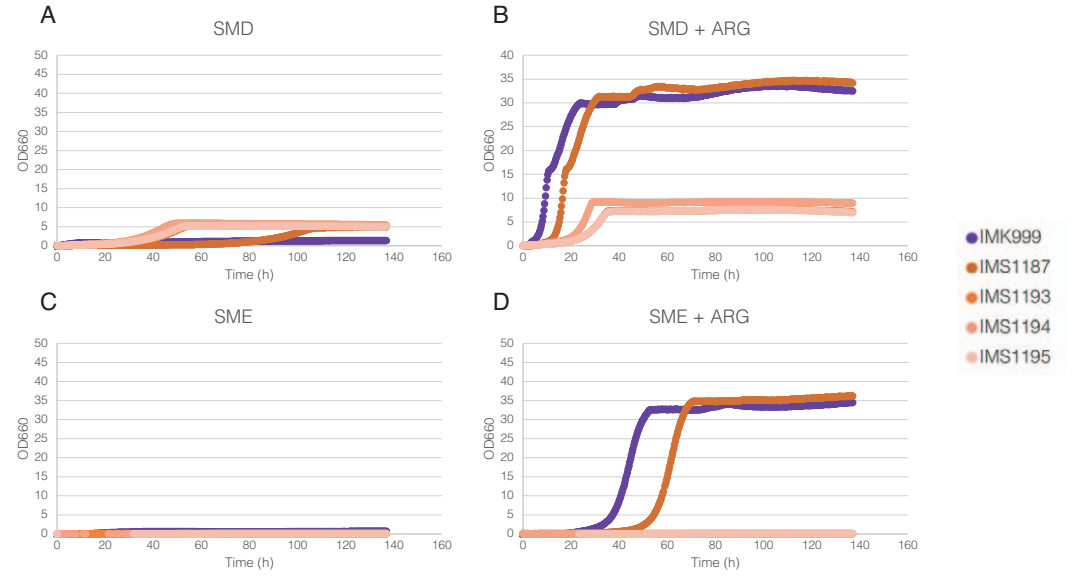


Figure S4.4. Growth of IMK999 (*Δarg8*, purple), the evolved arginine prototrophic population IMS1187 (dark orange) and its three derived single-colony isolated IMS1193, 1194 and 1195 (light orange). Growth was characterized in microtiter plates on SM-glucose (SMD, A), SMD + arginine (B), SM-ethanol (SME, C) and SME + arginine (D). The data represent the average and standard deviation of independent culture replicates.

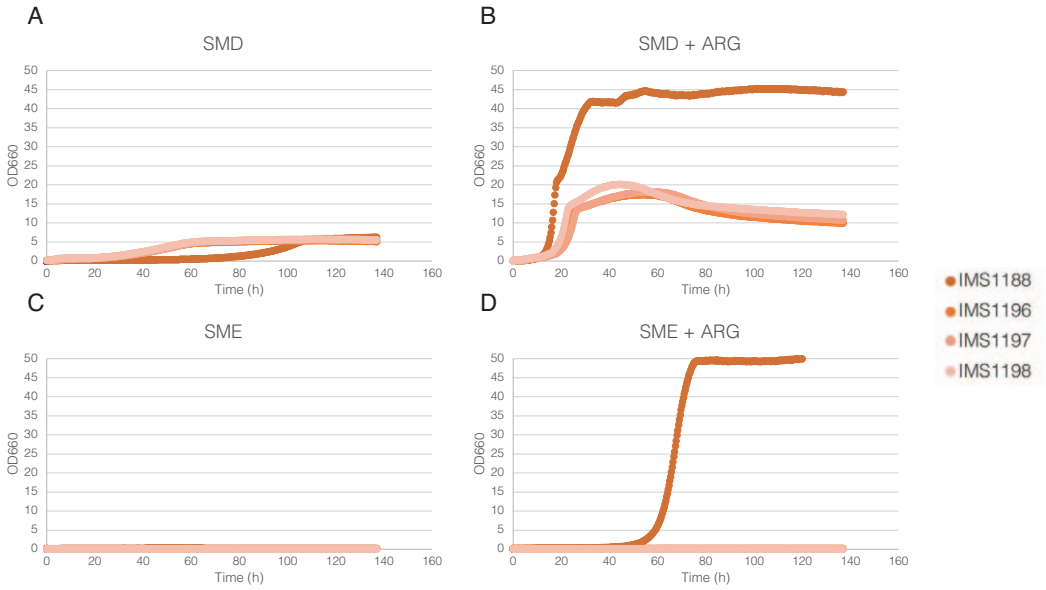


Figure S4.5. Growth of the evolved arginine prototrophic population IMS1188 (dark orange) and its three derived single-colony isolated IMS1196, 1197 and 1198 (light orange). Growth was characterized in microtiter plates on SM-glucose (SMD, A), SMD + arginine (B), SM-ethanol (SME, C) and SME + arginine (D). The data represent the average and standard deviation of independent culture replicates.

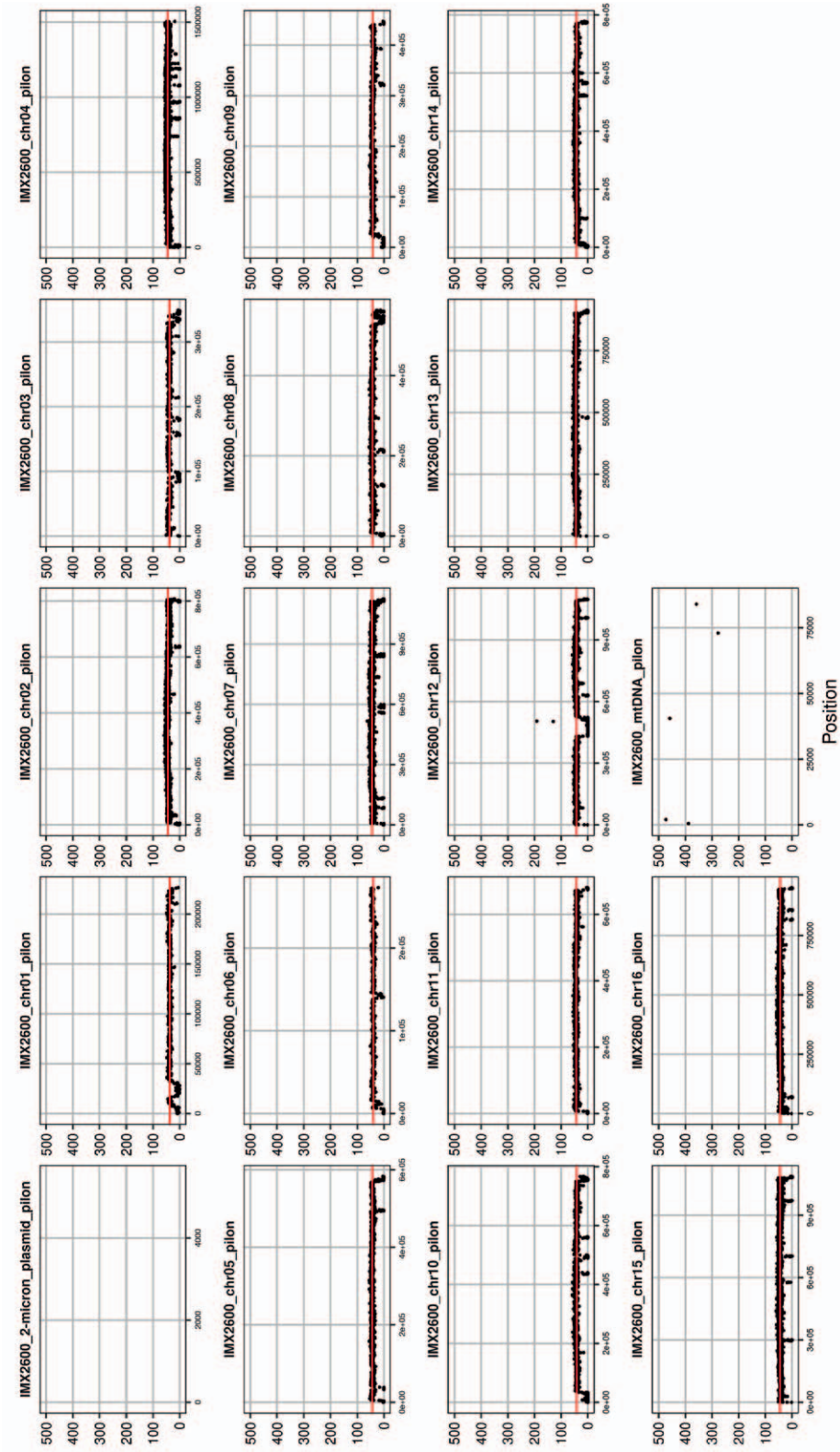


Figure S4.6. Sequencing coverage depth of the CEN.PK113-7D genome. Coverage depth was calculated as the average number of reads mapping to non-overlapping 500 bp sliding windows, obtained by short-read whole-genome sequencing of DNA, relative to the reference genome of IMX2600. The 2µ plasmid had a coverage >500 and is therefore not displayed.

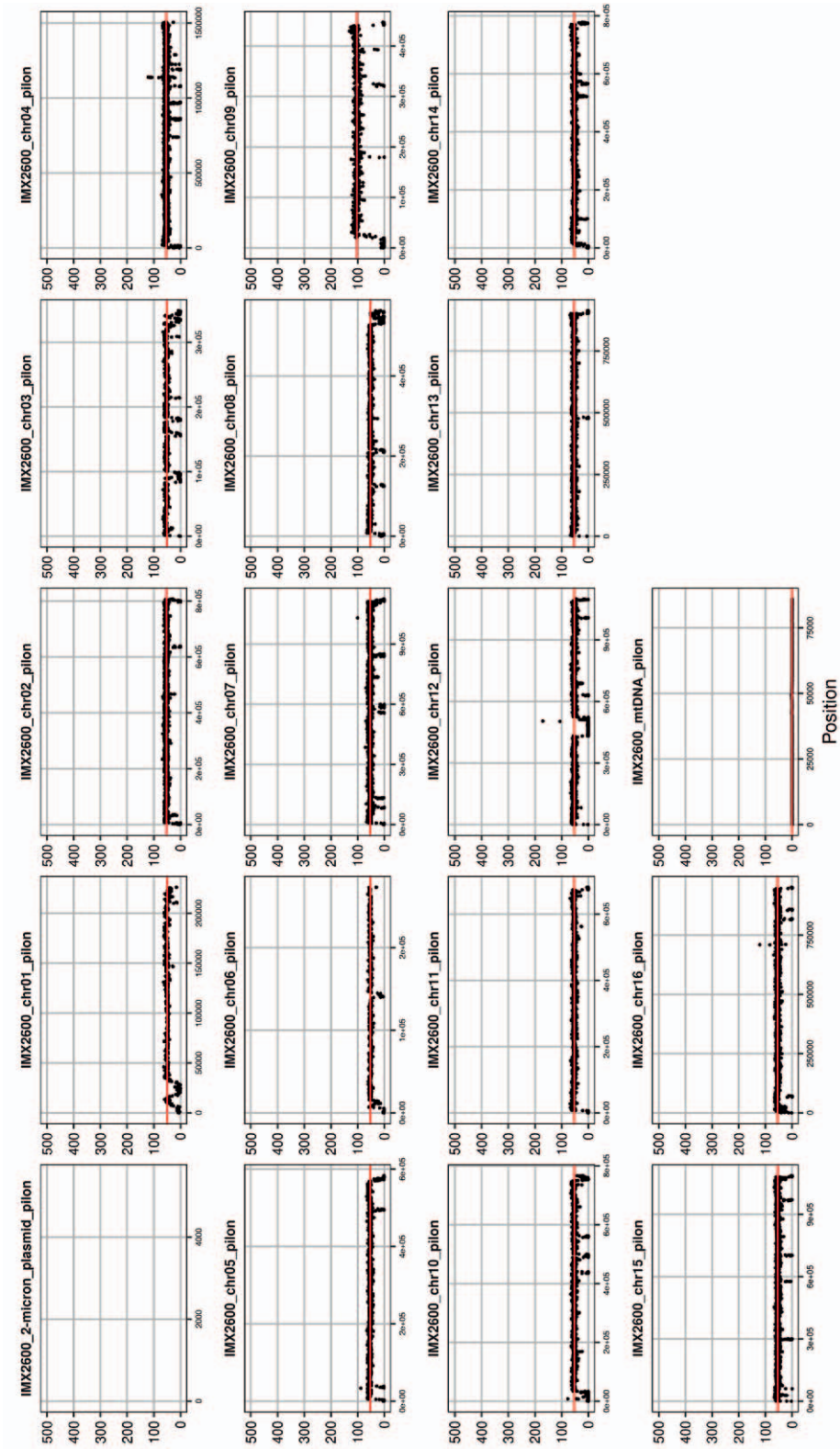


Figure S4.7. Sequencing coverage depth of the IMS1191 genome. Coverage depth was calculated as the average number of reads mapping to non-overlapping 500 bp sliding windows, obtained by short-read whole-genome sequencing of DNA, relative to the reference genome of IMX2600. The 2μ plasmid had a coverage >500 and is therefore not displayed

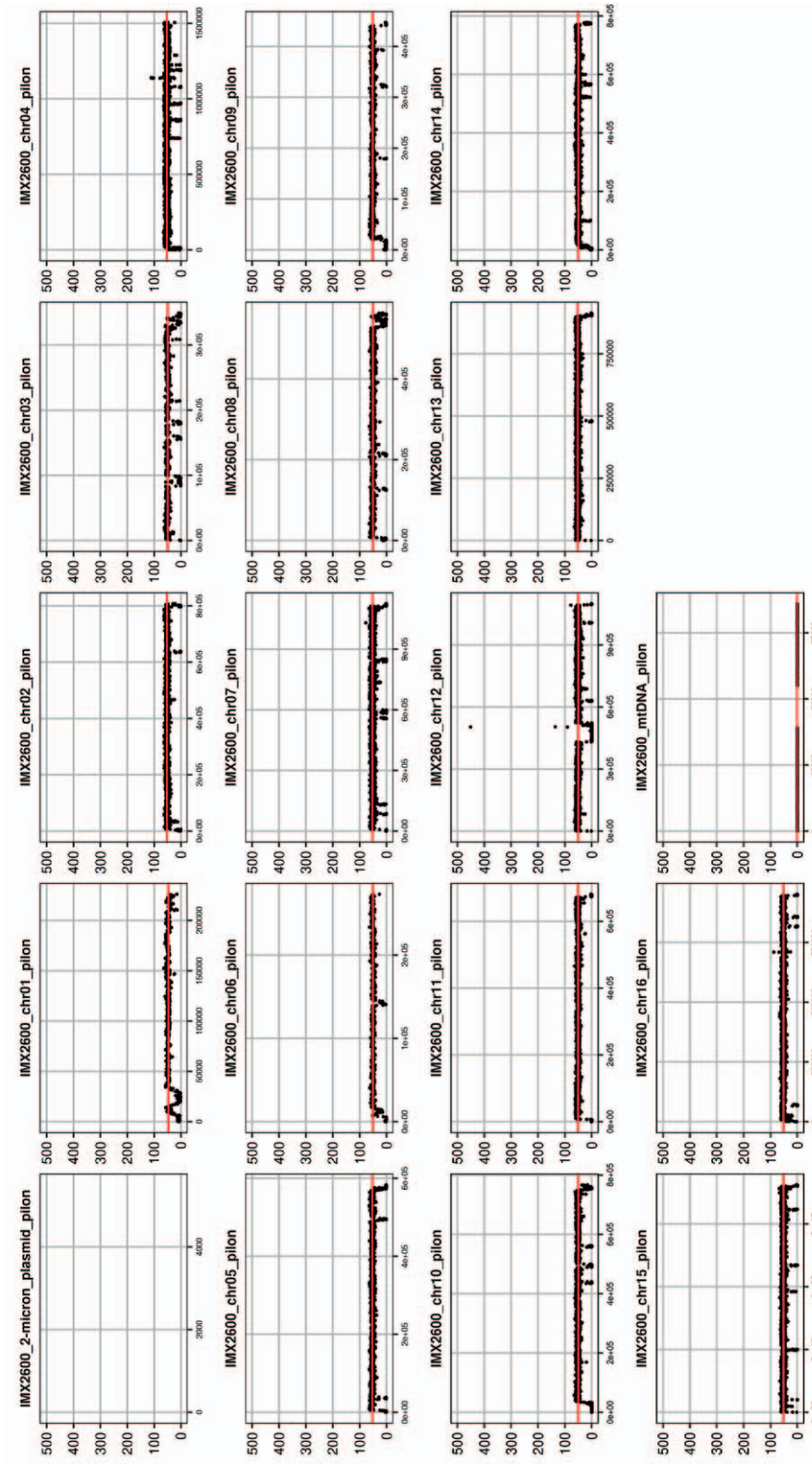


Figure S4.8. Sequencing coverage depth of the IMS1194 genome. Coverage depth was calculated as the average number of reads mapping to non-overlapping 500 bp sliding windows, obtained by short-read whole-genome sequencing of DNA, relative to the reference genome of IMX2600. The 2μ plasmid had a coverage >500 and is therefore not displayed

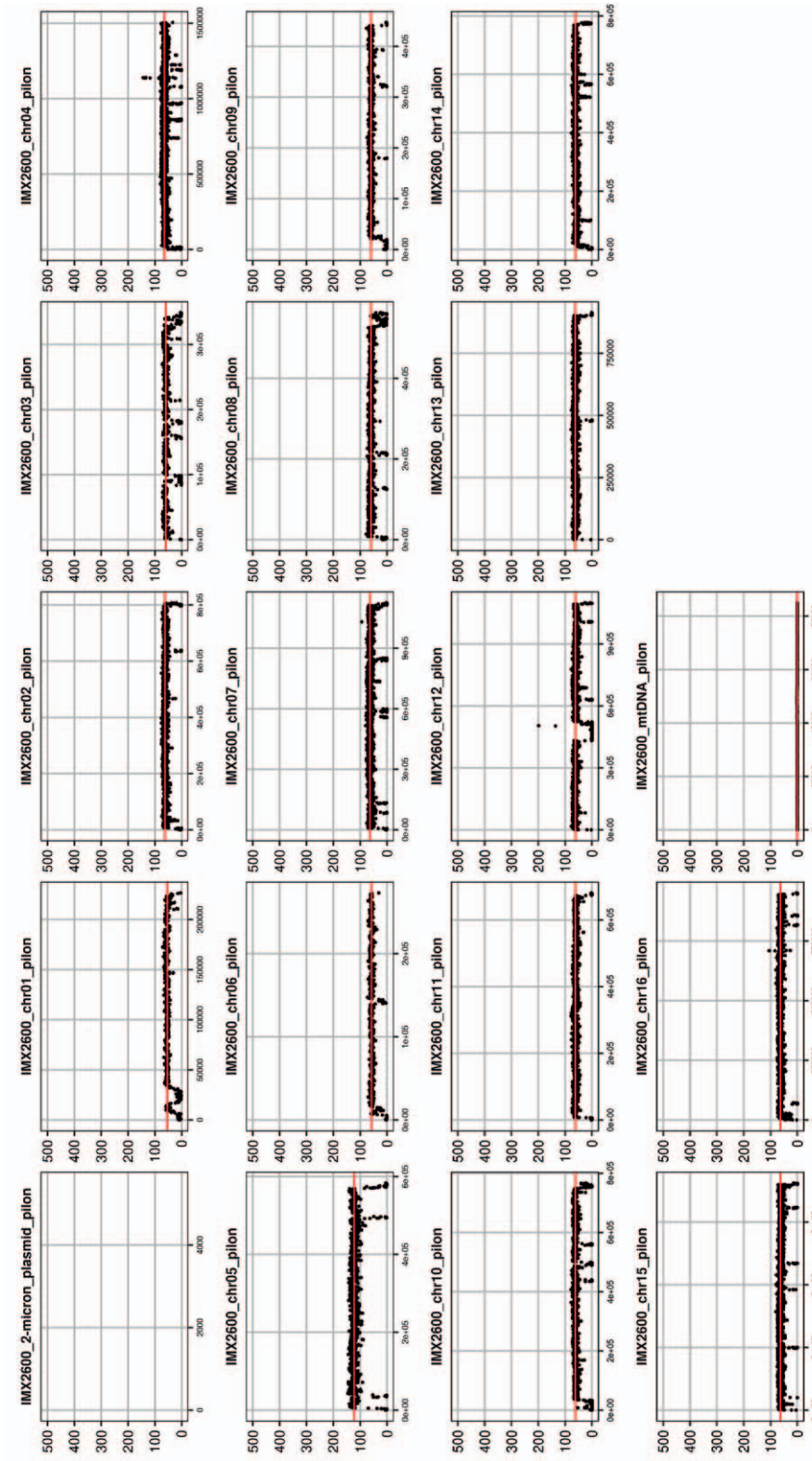


Figure S4.9. Sequencing coverage depth of the IMS1197 genome. Coverage depth was calculated as the average number of reads mapping to non-overlapping 500 bp sliding windows, obtained by short-read whole-genome sequencing of DNA, relative to the reference genome of IMX2600. The 2μ plasmid had a coverage >500 and is therefore not displayed

Figure S4.10. Alignment of *UME6* nucleotide sequence in the evolved and control strains. CLUSTAL O (version 1.2.4) multiple sequence alignment of subsets of the nucleotide sequences of *UME6* in CEN.PK113-7D, IMS1191, IMS1194 and IMS1197. Numbers indicate relative base pair positions in the coding sequence, starting at the start codon. Subsets of the genes are shown where insertions occurred. A) Base pair positions 540 – 720, the adenine deletion at position 650 in IMS1191 is highlighted in red. B) Base pair positions 2340– 2523, the adenine insertion at position 2509 in IMS1194/IMS1197 is highlighted in red.

```
UME6_CEN.PK113-7D       MLDKARSQSKHMDESNAAASLLSMETTANNHHYLHNKTSRATLMNSSQDGKKHAEDEVSD      60
UME6_IMS1194-1197       MLDKARSQSKHMDESNAAASLLSMETTANNHHYLHNKTSRATLMNSSQDGKKHAEDEVSD      60
UME6_IMS1191            MLDKARSQSKHMDESNAAASLLSMETTANNHHYLHNKTSRATLMNSSQDGKKHAEDEVSD      60

UME6_CEN.PK113-7D       GANSRHPTISSASIESLKTTYDENPLLSIMKSTCAPNNTPVHTPSGSPSLKVQSGGDIKD      120
UME6_IMS1194-1197       GANSRHPTISSASIESLKTTYDENPLLSIMKSTCAPNNTPVHTPSGSPSLKVQSGGDIKD      120
UME6_IMS1191            GANSRHPTISSASIESLKTTYDENPLLSIMKSTCAPNNTPVHTPSGSPSLKVQSGGDIKD      120

UME6_CEN.PK113-7D       DPKENDTTTTTNTTLQDRRDSDNAVHAAASPLAPSNTPSDPKSLCNGHVAQATDPQISGA      180
UME6_IMS1194-1197       DPKENDTTTTTNTTLQDRRDSDNAVHAAASPLAPSNTPSDPKSLCNGHVAQATDPQISGA      180
UME6_IMS1191            DPKENDTTTTTNTTLQDRRDSDNAVHAAASPLAPSNTPSDPKSLCNGHVAQATDPQISGA      180

UME6_CEN.PK113-7D       IQPQYTATNEDVFPYSSTSTNSNTATTTIVAGAKKKIHLPPPQAPAVSSPGTTA--AGSG      238
UME6_IMS1194-1197       IQPQYTATNEDVFPYSSTSTNSNTATTTIVAGAKKKIHLPPPQAPAVSSPGTTA--AGSG      238
UME6_IMS1191            IQPQYTATNEDVFPYSSTSTNSNTATTTIVAGAKKKNTFAATTSSSGFFPRYHRSRLGRG      240

UME6_CEN.PK113-7D       AGTGSGIRSRTGSDLPLIITSANKNNGKTTNSPMSILSRNNSTNNNDNNSIQSSDSRESS      298
UME6_IMS1194-1197       AGTGSGIRSRTGSDLPLIITSANKNNGKTTNSPMSILSRNNSTNNNDNNSIQSSDSRESS      298
UME6_IMS1191            HGLGHPFPHRIGFAAHHYQRQQEQR*                                 265

UME6_CEN.PK113-7D       NNNEIGGYLRGGTKRGGSPSNDSQVQHNVHDDQCAVGVAPRNFYFNKDREITDPNVKLDE      358
UME6_IMS1194-1197       NNNEIGGYLRGGTKRGGSPSNDSQVQHNVHDDQCAVGVAPRNFYFNKDREITDPNVKLDE      358

UME6_CEN.PK113-7D       NESKINISFWLNSKYRDEAYSLNESSSNNASSNTDTPTNSRHANTSSSITSRNNFQHFRF      418
UME6_IMS1194-1197       NESKINISFWLNSKYRDEAYSLNESSSNNASSNTDTPTNSRHANTSSSITSRNNFQHFRF      418

UME6_CEN.PK113-7D       NQIPSQPPTSASSFTSTNNNNPQRNNINRGEDPFATSSRPSTGFFYGDLPNRNNRNSPFH      478
UME6_IMS1194-1197       NQIPSQPPTSASSFTSTNNNNPQRNNINRGEDPFATSSRPSTGFFYGDLPNRNNRNSPFH      478

UME6_CEN.PK113-7D       TNEQYIPPPPPKYINSKLDGLRSRLLLGPNSASSSTKLDDDLGTAAAVLSNMRSSPYRTH      538
UME6_IMS1194-1197       TNEQYIPPPPPKYINSKLDGLRSRLLLGPNSASSSTKLDDDLGTAAAVLSNMRSSPYRTH      538

UME6_CEN.PK113-7D       DKPISNVNDMNNTNALGVPASRPHSSSFPSKGVLRPILLRIHNSEQQPIFESNNSTAVFD      598
UME6_IMS1194-1197       DKPISNVNDMNNTNALGVPASRPHSSSFPSKGVLRPILLRIHNSEQQPIFESNNSTAVFD      598

UME6_CEN.PK113-7D       EDQDQNQDLSPYHLNLNSKKVLDPTFESRTRQVTWNKNGKRIDRRLSAPEQQQQLEVPPL      658
UME6_IMS1194-1197       EDQDQNQDLSPYHLNLNSKKVLDPTFESRTRQVTWNKNGKRIDRRLSAPEQQQQLEVPPL      658

UME6_CEN.PK113-7D       KKSRRSVGNARVASQTNSDYNSLGESSTSSAPSSPSLKASSGLAYTADYPNATSPDFAKS      718
UME6_IMS1194-1197       KKSRRSVGNARVASQTNSDYNSLGESSTSSAPSSPSLKASSGLAYTADYPNATSPDFAKS      718

UME6_CEN.PK113-7D       KGKNVKPKAKSKAKQSSKKRPNNTTSKSKANNSQESNNATSSTSQGTRSRTGCWICRLRK      778
UME6_IMS1194-1197       KGKNVKPKAKSKAKQSSKKRPNNTTSKSKANNSQESNNATSSTSQGTRSRTGCWICRLRK      778

UME6_CEN.PK113-7D       KKCTEERPHCFNCERLKLDCHYDAFKPDFVSDPKKKQMKLEEIKKKTKEAKRRAMKKK*       836
UME6_IMS1194-1197       KKCTEERPHCFNCERLKLDCHYDAFKPDFVSDPKKKQMKLEEIKKKTKEAKRRAMKKKIK      838

UME6_IMS1194-1197       AH*     840
```

Figure S4.11 Alignment of Ume6 protein sequence in the evolved and control strains. CLUSTAL O (version 1.2.4) multiple sequence alignment of the protein products of *UME6* in CEN.PK113-7D (wildtype sequence), and mutants IMS1191 (adenine deletion at 650 bp, altered amino acid sequence highlighted in blue) and IMS11947/1197 (adenine insertion at 2509 bp, resulting amino acid extension highlighted in green).

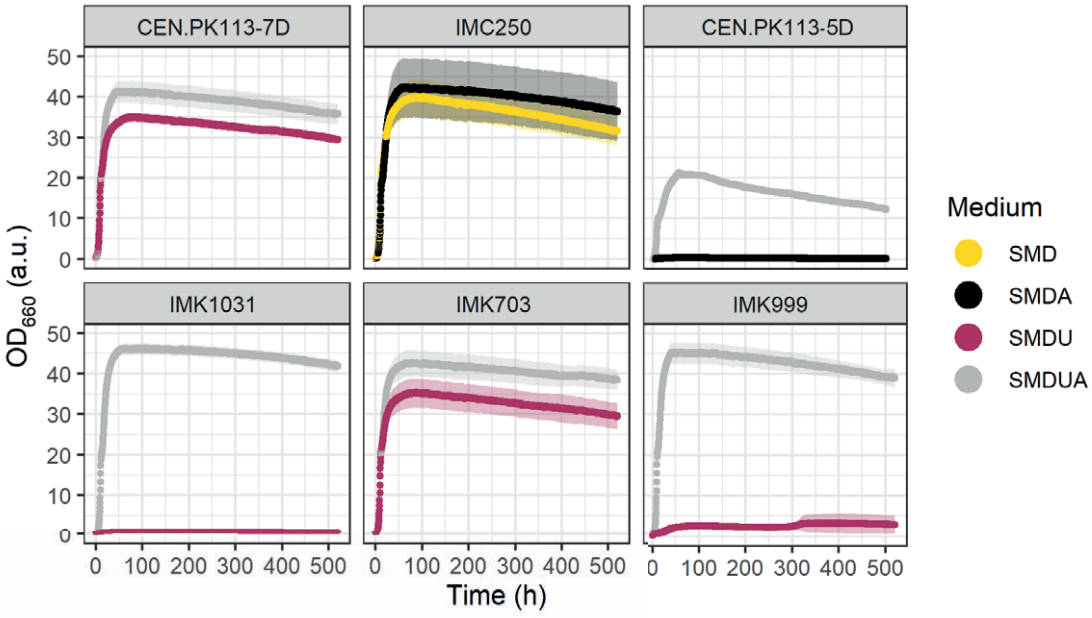


Figure S4.12. Growth of a mutant with double *ARG8* and *CAR2* deletion and its control strains in the presence and absence of arginine. Growth of strains CEN.PK113-7D (prototrophic), CEN.PK113-5D (ura3-52, uracil auxotroph), IMC250 *(Δarg8 Δcar2 Δura3*, plasmid-based expression of *preCOX4-ARG8*, *URA3*), IMK1031 *(Δarg8 Δcar2 Δura3*), IMK703 *(Δcar2*) and IMK999 (*Δarg8*) in SMD (yellow) SMD + arginine (SMDA, black), SMD + uracil (SMDU, maroon) and SMDU + arginine (SMDUA, grey). Data represent the average of biological triplicates in microtiter plates, the standard deviation is shown as a shaded area.

Table S4.1. Primers used in this study

# ID	Purpose	Sequence[1]
13807	*URA3* repair	CGGTTTCCTTGAAATTTTTTTGATTCGGTAATCTCCGAACAGAAGGAAGAACGAAGGAAGGGAATCTCGGTCGTAATGATTTCTATAATGACGAAAAAAAAAAAATTGGAAAGAAAAAGC
13808		GCTTTTTCTTTCCAATTTTTTTTTTTTTCGTCATTATAGAAATCATTACGACCGAGATTCCCTTCCTTCGTTCTTCCTTCTGTTCGGAGATTACCGAATCAAAAAAATTTCAAGGAAACCG
6005	pMEL/pROS linearization	GATCATTTATCTTTCACTGCGGAGAAG
6006		GTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTC
18329	*ARG2* gRNA	TGCGCATGTTTCGGCGTTCGAAACTTCTCCGCAGTGAAAGATAAATGATCCAAAATATACAAATGATAGTGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAAC
18330		GTTGATAACGGACTAGCCTTATTTTAACTTGCTATTTCTAGCTCTAAAACACTATCATTTGTATATTTTGGATCATTTATCTTTCACTGCGGAGAAGTTTCGAACGCCGAAACATGCGCA
18331	Δ*arg2* repair	AAAGTTAACACATTTGAAGCTAGGACCGTCAATCTAGGACACTAGTCGGTTACTTTTATCAACAAAGACGCATTCTCCTCCTCCTTAACTTTCAAGTTTTCATTTTCAGTTTTTAAGGTA
18332		TACCTTAAAAACTGAAAATGAAAACTTGAAAGTTAAGGAGGAGGAGAATGCGTCTTTGTTGATAAAAGTAACCGACTAGTGTCCTAGATTGACGGTCCTAGCTTCAAATGTGTTAACTTT
18333	(Δ)*ARG2* colony PCR	CTTCCCATGGAGTCATAAGC
18334		CGCCTTACTTTGAGGAGAAC
18335	*ARG8* gRNA	TGCGCATGTTTCGGCGTTCGAAACTTCTCCGCAGTGAAAGATAAATGATCTTGAAGTTTGCAAAGAAACAGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAAC
18336		GTTGATAACGGACTAGCCTTATTTTAACTTGCTATTTCTAGCTCTAAAACTGTTTCTTTGCAAACTTCAAGATCATTTATCTTTCACTGCGGAGAAGTTTCGAACGCCGAAACATGCGCA
18337	Δ*arg8* repair	TGCAAGTGACTGCGAACATTTTTTTCGTTTGTTAGAATAATTCAAGAATCGCTACCAATCATATACTTGTAAATTGTCATGTATAGATTGTTTTTTTTTTTTTCATCTTTATATTGTAG
18338		CTACAATATAAAGATGAAAAAAAAAAAAACAATCTATACATGACAATTTACAAAGTATATGATTGGTAGCGATTCTTGAATTATTCTAACAAACGAAAAAAATGTTCGCAGTCACTTGCA
18339	(Δ)*ARG8* colony PCR	AATGCTCCTATCGGGATTCG
18340		ATGTGTCTCTGACCGAATGG
8818	*CAR2* gRNA 1	TGCGCATGTTTCGGCGTTCGAAACTTCTCCGCAGTGAAAGATAAATGATCGAAGAATTCGGTCACGAACTGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAAC
8819	*CAR2* gRNA 2	TGCGCATGTTTCGGCGTTCGAAACTTCTCCGCAGTGAAAGATAAATGATCGTCGCGGATGACCTCGAGGGGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAAC
8814	Δ*car2* repair	GTATTACAGCAAAGGATACTACACCACAAGACACCACAGCAGCACAGAACCACAACACCATATTAATACGCATTTTTTTTATGTTTTCTCGTTTCTTTTCTTTTTTTTGTTTTATATCTT
8815		AAGATATAAAACAAAAAAAAGAAAAGAAACGAGAAAACATAAAAAAAATGCGTATTAATATGGTGTTGTGGTTCTGTGCTGCTGTGGTGTCTTGTGGTGTAGTATCCTTTGCTGTAATAC
8816	(Δ)*CAR2* colony PCR	CTGCTTGCCCATTCCAGTTGTGC
8817		TCTTCTTCCTGATGCGCCC
19477	*ARG8-MTS*, part 3b	GCATCGTCTCATC**GGTCTC**ATTCTGCCTTTCAAGTGACCACTTACTCTAGACC
19478		ATGCCGTCTCA**GGTCTC**AGGATTTAAGCGTAAACCGCTTCAATAGCC
17711	*preCOX4-MTS* part 3a	GCATCGTCTCATC**GGTCTC**ATATGTTGTCTTTG
17712		ATGCCGTCTCA**GGTCTC**AAGAACCCAACAAG
18445	Confirmation pUDC434	AAAGCACCCATTGTTCTACC
13048		ATAGCTTCCGGCACATGGTGAGATGCACTCTAACCGTCAGGCGACGTATAGATTCGTATGGCCGATAATTGCAGACGAAC

[1] Underlined = gRNA sequences, **bold** = BsaI sites

Table S4.2. Plasmids used in this study

Plasmid name	*Relevant genotype*	*Purpose*	*Source*
pROS10	*tSUP4 URA3, 2μ, bla pSNR52*	Backbone for Cas9-mediated cleavage	[45]
pMEL13	*tSUP4 KanMX, 2μ, bla pSNR52*	Backbone for Cas9-mediated cleavage	[45]
pUDR107	*pSNR52-URA3-gRNA-tSUP4, hphNT1, 2μ, bla*	Cas9-mediated cleavage of URA3	[86]
pUDR131	*pSNR52-CAR2-gRNA-tSUP4, hphNT1, 2μ, bla*	Cas9-mediated cleavage of CAR2	This study
pYTK001	*YTK entry vector, CamR*	YTK part cloning	[46]
pYTK009	*pTDH3, CamR*	YTK part 2 pTDH3	[46]
pGGKp321	*preCOX4-MTS, CamR*	YTK part 3a preCOX4-MTS	This study
pYTK053	*tADH1, CamR*	YTK part 2 tADH1	[46]
pUD538	*URA3, CEN6/ARS4, bla, ColE1*	YTK drop-out backbone	[50]
pUDC286	*(preCOX4-MTS)-mRuby2*	Template for preCOX4-MTS	[50]
pUDC434	*pTDH3-preCOX4MTS-ARG8-tADH1, URA3, CEN6/ARS4, Amp, ColE1*	Expression and mitochondrial targeting of preCOX4-Arg8p	This study

References

1. Malina, C., Larsson, C., and Nielsen, J., Yeast mitochondria: an overview of mitochondrial biology and the potential of mitochondrial systems biology. *FEMS Yeast Research*, *2018*. ***18****(5): p. foy040*.
2. Braun, R.J. and Westermann, B., Mitochondrial dynamics in yeast cell death and aging. *Biochemical Society Transactions*, *2011*. ***39****(5): p. 1520-1526*.
3. Baile, M.G. and Claypool, S.M., The power of yeast to model diseases of the powerhouse of the cell. *Frontiers in bioscience (Landmark edition)*, *2013*. ***18****: p. 241*.
4. Contamine, V. and Picard, M., Maintenance and integrity of the mitochondrial genome: A plethora of nuclear genes in the budding yeast. *Microbiology and Molecular Biology Reviews*, *2000*. ***64****(2): p. 281-315*.
5. Larosa, V. and Remacle, C., Transformation of the mitochondrial genome. *International Journal of Developmental Biology*, *2013*. ***57****(6-7-8): p. 659-665*.
6. Bonnefoy, N. and Fox, T.D., Directed alteration of *Saccharomyces cerevisiae* mitochondrial DNA by biolistic transformation and homologous recombination, in mitochondria: Practical protocols, D. Leister and J.M. Herrmann, Editors. 2007, *Humana Press*: Totowa, NJ. p. 153-166.
7. Gammage, P.A., Moraes, C.T., and Minczuk, M., Mitochondrial genome engineering: the revolution may not be CRISPR-Ized. *Trends in Genetics*, *2018*. ***34****(2): p. 101-110*.
8. Yoo, B.-C., Yadav, N.S., Orozco Jr, E.M., and Sakai, H., Cas9/gRNA-mediated genome editing of yeast mitochondria and *Chlamydomonas* chloroplasts. *PeerJ*, *2020*. ***8****: p. e8362*.
9. Loutre, R., Heckel, A.M., Smirnova, A., Entelis, N., and Tarassov, I., Can Mitochondrial DNA be CRISPRized: *Pro and Contra*. *IUBMB life*, *2018*. ***70****(12): p. 1233-1239*.
10. Jo, A., Ham, S., Lee, G.H., Lee, Y.I., Kim, S., Lee, Y.S., Shin, J.H., and Lee, Y., Efficient mitochondrial genome editing by CRISPR/Cas9. *Biomed Res Int*, *2015*. ***2015****: p. 305716*.
11. Amai, T., Tsuji, T., Ueda, M., and Kuroda, K., Development of a mito-CRISPR system for generating mitochondrial DNA-deleted strain in *Saccharomyces cerevisiae*. *Bioscience, Biotechnology, and Biochemistry*, *2020*. ***85****(4): p. 895-901*.
12. Dong, C., Shi, Z., Huang, L., Zhao, H., Xu, Z., and Lian, J., Cloning and characterization of a panel of mitochondrial targeting sequences for compartmentalization engineering in *Saccharomyces cerevisiae*. *Biotechnology and bioengineering*, *2021*.
13. Malina, C., Larsson, C., and Nielsen, J., Yeast mitochondria: an overview of mitochondrial biology and the potential of mitochondrial systems biology. *FEMS Yeast Res*, *2018*. ***18****(5)*.
14. Kolesnikova, O., Kazakova, H., Comte, C., Steinberg, S., Kamenski, P., Martin, R.P., Tarassov, I., and Entelis, N., Selection of RNA aptamers imported into yeast and human mitochondria. *RNA*, *2010*. ***16****(5): p. 926-41*.
15. Wang, G., Chen, H.-W., Oktay, Y., Zhang, J., Allen, E.L., Smith, G.M., Fan, K.C., Hong, J.S., French, S.W., and McCaffery, J.M.J.C., PNPASE regulates RNA import into mitochondria. *2010*. ***142****(3): p. 456-467*.
16. Weber-Lotfi, F. and Dietrich, A., Targeting therapeutic nucleic acids into mitochondria: A long challenge, in Mitochondrial biology and experimental therapeutics. 2018, *Springer*. p. 565-592.
17. Schmiderer, L., Yudovich, D., Oburoglu, L., Hjort, M., and Larsson, J., Site-specific CRISPR-based mitochondrial DNA manipulation is limited by gRNA import. *Scientific reports*, *2022*. ***12****(1): p. 18687*.
18. Yuan, P., Mao, X., Wu, X., Liew, S.S., Li, L., and Yao, S.Q., Mitochondria-targeting, intracellular delivery of native proteins using biodegradable silica nanoparticles. *Angewandte Chemie International Edition*, *2019*. ***58****(23): p. 7657-7661*.
19. Verechshagina, N., Nikitchina, N., Yamada, Y., Harashima, H., Tanaka, M., Orishchenko, K., and Mazunin, I., Future of human mitochondrial DNA editing technologies. *Mitochondrial DNA Part A*, *2018: p. 1-8*.
20. Weber-Lotfi, F., Ibrahim, N., Boesch, P., Cosset, A., Konstantinov, Y., Lightowlers, R.N., and Dietrich, A., Developing a genetic approach to investigate the mechanism of mitochondrial competence for DNA import. *Biochimica et Biophysica Acta (BBA)-Bioenergetics*, *2009*. ***1787****(5): p. 320-327*.
21. Suhm, T., Habernig, L., Rzepka, M., Kaimal, J.M., Andreasson, C., Buttner, S., and Ott, M., A novel system to monitor mitochondrial translation in yeast. *Microbial Cell*, *2018*. ***5****(3): p. 158-164*.
22. Mireau, H., Arnal, N., and Fox, T.D., Expression of Barstar as a selectable marker in yeast mitochondria. *Molecular Genetics & Genomics*, *2003*. ***270****(1): p. 1-8*.
23. Jauniaux, J.C., Urrestarazu, L.A., and Wiame, J.M., Arginine metabolism in *Saccharomyces cerevisiae*: subcellular localization of the enzymes. *Journal of bacteriology*, *1978*. ***133****(3): p. 1096-1107*.
24. Fox, T.D., Natural variation in the genetic code. *Annual review of genetics*, *1987*. ***21****(1): p. 67-91*.
25. Steele, D.F., Butler, C.A., and Fox, T.D., Expression of a recoded nuclear gene inserted into yeast mitochondrial DNA is limited by mRNA-specific translational activation. *Proceedings of the National Academy of Sciences*, *1996*. ***93****(11): p. 5253-5257*.
26. Stenger, M., Le, D.T., Klecker, T., and Westermann, B., Systematic analysis of nuclear gene function in respiratory growth and expression of the mitochondrial genome in *S. cerevisiae*. *Microbial Cell*, *2020*. ***7****(9): p. 234-249*.
27. Barros, M.H. and Tzagoloff, A., Aep3p-dependent translation of yeast mitochondrial *ATP8*. *Molecular Biology of the Cell*, *2017*. ***28****(11): p. 1426-1434*.
28. Bonnefoy, N. and Fox, T.D., *In vivo* analysis of mutated initiation codons in the mitochondrial *COX2* gene of *Saccharomyces cerevisiae* fused to the reporter gene *ARG8* reveals lack of downstream reinitiation. *Molecular and General Genetics MGG*, *2000*. ***262****(6): p. 1036-1046*.
29. Zhang, H., Chatterjee, A., and Singh, K.K., *Saccharomyces cerevisiae* Polymerase ζ Functions in mitochondria. *Genetics*, *2006*. ***172****(4): p. 2683-2688*.
30. Rak, M., Tetaud, E., Godard, F., Sagot, I., Salin, B., Duvezin-Caubet, S., Slonimski, P.P., Rytka, J., and di Rago, J.-P., Yeast cells lacking the mitochondrial gene encoding the ATP synthase subunit 6 exhibit a selective loss of complex iv and unusual mitochondrial morphology *Journal of Biological Chemistry*, *2007*. ***282****(15): p. 10853-10864*.
31. Perez-Martinez, X., Broadley, S.A., and Fox, T.D., Mss51p promotes mitochondrial Cox1p synthesis and interacts with newly synthesized Cox1p. *The EMBO journal*, *2003*. ***22****(21): p. 5951-5961*.
32. De Silva, D., Poliquin, S., Zeng, R., Zamudio-Ochoa, A., Marrero, N., Perez-Martinez, X., Fontanesi, F., and Barrientos, A., The DEAD-box helicase Mss116 plays distinct roles in mitochondrial ribogenesis and mRNA-specific translation. *Nucleic acids research*, *2017*. ***45****(11): p. 6628-6643*.
33. Watts, T., Khalimonchuk, O., Wolf, R.Z., Turk, E.M., Mohr, G., and Winge, D.R., Mne1 is a novel component of the mitochondrial splicing apparatus responsible for processing of a *COX1* group I intron in yeast *Journal of Biological Chemistry*, *2011*. ***286****(12): p. 10137-10146*.
34. Williams, E.H. and Fox, T.D., Antagonistic signals within the *COX2* mRNA coding sequence control its translation in *Saccharomyces cerevisiae* mitochondria. *RNA*, *2003*. ***9****(4): p. 419-431*.
35. Entian, K.-D. and Kötter, P., 25 Yeast genetic strain and plasmid collections, in Methods in Microbiology, I. Stansfield and M.J.R. Stark, Editors. 2007, *Academic Press*. p. 629-666.
36. Baker Brachmann, C., Davies, A., Cost, G.J., Caputo, E., Li, J., Hieter, P., and Boeke, J.D., Designer deletion strains

derived from *Saccharomyces cerevisiae* S288C: a useful set of strains and plasmids for PCR-mediated gene disruption and other applications. *Yeast*, *1998*. ***14****(2): p. 115-132*.

37. Giaever, G., Chu, A.M., Ni, L., Connelly, C., Riles, L., Véronneau, S., Dow, S., Lucau-Danila, A., Anderson, K., André, B.,... Johnston, M., Functional profiling of the *Saccharomyces cerevisiae* genome. *Nature*, *2002*. ***418****(6896): p. 387-391*.
38. Verduyn, C., Postma, E., Scheffers, W.A., and Van Dijken, J.P., Effect of benzoic acid on metabolic fluxes in yeasts: A continuous-culture study on the regulation of respiration and alcoholic fermentation. *Yeast*, *1992*. ***8****(7): p. 501-517*.
39. Pronk, J.T., Auxotrophic yeast strains in fundamental and applied research. *Applied and environmental microbiology*, *2002*. ***68****(5): p. 2095-2100*.
40. Oura, E., Effect of aeration on the growth energetics and biochemical composition of baker's yeast. 1972, Research Laboratories of the State Alcohol Monopoly.
41. Mans, R., van Rossum, H.M., Wijsman, M., Backx, A., Kuijpers, N.G., van den Broek, M., Daran-Lapujade, P., Pronk, J.T., van Maris, A.J., and Daran, J.-M.G., CRISPR/Cas9: a molecular Swiss army knife for simultaneous introduction of multiple genetic modifications in *Saccharomyces cerevisiae*. *FEMS yeast research*, *2015*. ***15****(2)*.
42. Randazzo, P., Bennis, N.X., Daran, J.-M., and Daran-Lapujade, P., gEL DNA: A cloning- and polymerase chain reaction–free method for CRISPR-based multiplexed genome editing. *The CRISPR Journal*, *2021*. ***4****(6): p. 896-913*.
43. Koster, C.C., Kleefeldt, A., van den Broek, M., Luttik, M., Daran, J.-M., and Daran-Lapujade, P., Long-read direct RNA sequencing of the mitochondrial transcriptome of *Saccharomyces cerevisiae* reveals condition-dependent intron turnover. *bioRxiv*, *2023: p. 2023.01. 19.524680*.
44. Wronska, A.K., van den Broek, M., Perli, T., de Hulster, E., Pronk, J.T., and Daran, J.-M., Engineering oxygen-independent biotin biosynthesis in *Saccharomyces cerevisiae*. *Metabolic Engineering*, *2021*. ***67****: p. 88-103*.
45. Mans, R., Wijsman, M., Daran-Lapujade, P., and Daran, J.-M., A protocol for introduction of multiple genetic modifications in *Saccharomyces cerevisiae* using CRISPR/Cas9. *FEMS yeast research*, *2018*. ***18****(7): p. foy063*.
46. Lee, M.E., DeLoache, W.C., Cervantes, B., and Dueber, J.E., A highly characterized yeast toolkit for modular, multipart assembly. *ACS synthetic biology*, *2015*. ***4****(9): p. 975-986*.
47. Vowinckel, J., Hartl, J., Butler, R., and Ralser, M., MitoLoc: A method for the simultaneous quantification of mitochondrial network morphology and membrane potential in single cells. *Mitochondrion*, *2015*. ***24****: p. 77-86*.
48. Gietz, R.D. and Woods, R.A., Transformation of yeast by lithium acetate/single-stranded carrier DNA/polyethylene glycol method, in Methods in enzymology. 2002, *Elsevier*. p. 87-96.
49. Boonekamp, F.J., Knibbe, E., Vieira-Lara, M.A., Wijsman, M., Luttik, M.A.H., van Eunen, K., den Ridder, M., Bron, R., Almonacid Suarez, A.M., van Rijn, P.,... Daran-Lapujade, P., Full humanization of the glycolytic pathway in *Saccharomyces cerevisiae*. *Cell reports*, *2022*. ***39****(13): p. 111010*.
50. Bouwknegt, J., Koster, C.C., Vos, A.M., Ortiz-Merino, R.A., Wassink, M., Luttik, M.A.H., van den Broek, M., Hagedoorn, P.L., and Pronk, J.T., Class-II dihydroorotate dehydrogenases from three phylogenetically distant fungi support anaerobic pyrimidine biosynthesis. *Fungal Biology and Biotechnology*, *2021*. ***8****(1): p. 10*.
51. Schindelin, J., Arganda-Carreras, I., Frise, E., Kaynig, V., Longair, M., Pietzsch, T., Preibisch, S., Rueden, C., Saalfeld, S., Schmid, B.,... Cardona, A., Fiji: an open-source platform for biological-image analysis. *Nature methods*, *2012*. ***9****(7): p. 676-682*.
52. Li, H. and Durbin, R., Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics*, *2009*. ***25****(14): p. 1754-60*.
53. Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., and Durbin, R., The Sequence Alignment/Map format and SAMtools. *Bioinformatics*, *2009*. ***25****(16): p. 2078-9*.
54. Walker, B.J., Abeel, T., Shea, T., Priest, M., Abouelliel, A., Sakthikumar, S., Cuomo, C.A., Zeng, Q., Wortman, J., Young, S.K., and Earl, A.M., Pilon: An integrated tool for comprehensive microbial variant detection and genome assembly improvement. *PloS one*, *2014*. ***9****(11): p. e112963*.
55. Robinson, J.T., Thorvaldsdóttir, H., Winckler, W., Guttman, M., Lander, E.S., Getz, G., and Mesirov, J.P., Integrative genomics viewer. *Nature biotechnology*, *2011*. ***29****(1): p. 24-6*.
56. Nijkamp, J., van den Broek, M., Datema, E., de Kok, S., Bosman, L., Luttik, M., Pronk, J., de Ridder, D., and Daran, J.-M., *De novo* sequencing, assembly and analysis of the genome of the laboratory strain *Saccharomyces cerevisiae* CEN. PK113-7D, a model for modern industrial biotechnology. *Microbial cell factories*, *2012*. ***11****: p. 36*.
57. Sievers, F., Wilm, A., Dineen, D., Gibson, T.J., Karplus, K., Li, W., Lopez, R., McWilliam, H., Remmert, M., Söding, J., Thompson, J.D., and Higgins, D.G., Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Molecular systems biology*, *2011*. ***7****(1): p. 539*.
58. Al-Saryi, N.A., Al-Hejjaj, M.Y., van Roermund, C.W.T., Hulmes, G.E., Ekal, L., Payton, C., Wanders, R.J.A., and Hettema, E.H., Two NAD-linked redox shuttles maintain the peroxisomal redox balance in *Saccharomyces cerevisiae*. *Scientific reports*, *2017*. ***7****(1): p. 11868*.
59. Hammer, S.K. and Avalos, J.L., Uncovering the role of branched-chain amino acid transaminases in *Saccharomyces cerevisiae* isobutanol biosynthesis. *Metabolic Engineering*, *2017*. ***44****: p. 302-312*.
60. Cherry, J.M., Hong, E.L., Amundsen, C., Balakrishnan, R., Binkley, G., Chan, E.T., Christie, K.R., Costanzo, M.C., Dwight, S.S., Engel, S.R.,... Wong, E.D., *Saccharomyces* Genome Database: the genomics resource of budding yeast. *Nucleic acids research*, *2012*. ***40****(Database issue): p. D700-5*.
61. Vandenbol, M. and Portetelle, D., Disruption of six ORFs on *Saccharomyces cerevisiae* chromosome X: the YJL069c gene of unknown function is essential to cell viability. *Yeast*, *1999*. ***15****(13): p. 1411-1417*.
62. Abadjieva, A., Pauwels, K., Hilven, P., and Crabeel, M., A new yeast metabolon involving at least the two first enzymes of arginine biosynthesis: acetylglutamate synthase activity requires complex formation with acetylglutamate kinase. *Journal of Biological Chemistry*, *2001*. ***276****(46): p. 42869-80*.
63. El Alami, M., Dubois, E., Oudjama, Y., Tricot, C., Wouters, J., Stalon, V., and Messenguy, F., Yeast epiarginase regulation, an enzyme-enzyme activity control: Identification of residues of ornithine carbamoyltransferase and arginase responsible for enzyme catalytic and regulatory activities *Journal of Biological Chemistry*, *2003*. ***278****(24): p. 21550-21558*.
64. Pande, S., Merker, H., Bohl, K., Reichelt, M., Schuster, S., de Figueiredo, L.F., Kaleta, C., and Kost, C., Fitness and stability of obligate cross-feeding interactions that emerge upon gene loss in bacteria. *The ISME Journal*, *2014*. ***8****(5): p. 953-962*.
65. Bachmann, H., Pronk, J.T., Kleerebezem, M., and Teusink, B., Evolutionary engineering to enhance starter culture performance in food fermentations. *Current opinion in biotechnology*, *2015*. ***32****: p. 1-7*.
66. Foury, F., Roganti, T., Lecrenier, N., and Purnelle, B., The complete sequence of the mitochondrial genome of *Saccharomyces cerevisiae*. *FEBS letters*, *1998*. ***440****(3): p. 325-331*.
67. Washburn, B.K. and Esposito, R.E., Identification of the Sin3-binding site in Ume6 defines a two-step process for conversion of Ume6 from a transcriptional repressor to an activator in yeast. *Molecular and Cellular Biology*, *2001*. ***21****(6): p. 2057-2069*.
68. Lardenois, A., Becker, E., Walther, T., Law, M.J., Xie, B., Demougin, P., Strich, R., and Primig, M., Global alterations of the transcriptional landscape during yeast growth and

development in the absence of Ume6-dependent chromatin modification. *Molecular Genetics and Genomics*, *2015*. ***290****(5): p. 2031-2046*.

69. Williams, R.M., Primig, M., Washburn, B.K., Winzeler, E.A., Bellis, M., Sarrauste de Menthière, C., Davis, R.W., and Esposito, R.E., The Ume6 regulon coordinates metabolic and meiotic gene expression in yeast. *Proceedings of the National Academy of Sciences*, *2002*. ***99****(21): p. 13431-13436*.
70. Kadosh, D. and Struhl, K., Repression by Ume6 involves recruitment of a complex containing Sin3 corepressor and Rpd3 histone deacetylase to target promoters. *Cell*, *1997*. ***89****(3): p. 365-371*.
71. Messenguy, F., Vierendeels, F., Scherens, B., and Dubois, E., In *Saccharomyces cerevisiae*, expression of arginine catabolic genes *CAR1* and *CAR2* in response to exogenous nitrogen availability is mediated by the Ume6 (CargRI)-Sin3 (CargRII)-Rpd3 (CargRIII) complex. *Journal of bacteriology*, *2000*. ***182****(11): p. 3158-3164*.
72. Sweet, D.H., Jang, Y.K., and Sancar, G.B., Role of *UME6* in transcriptional regulation of a DNA repair gene in *Saccharomyces cerevisiae*. *Molecular and Cellular Biology*, *1997*. ***17****(11): p. 6223-6235*.
73. Strich, R., Surosky, R.T., Steber, C., Dubois, E., Messenguy, F., and Esposito, R.E., *UME6* is a key regulator of nitrogen repression and meiotic development. *Genes & development*, *1994*. ***8****(7): p. 796-810*.
74. Anderson, S.F., Steber, C.M., Esposito, R.E., and Coleman, J.E., *UME6*, a negative regulator of meiosis in *Saccharomyces cerevisiae*, contains a C-terminal Zn2Cys6 binuclear cluster that binds the URS1 DNA sequence in a zinc-dependent manner. *Protein Science*, *1995*. ***4****(9): p. 1832-1843*.
75. Donovan, D.A., Crandall, J.G., Truong, V.N., Vaaler, A.L., Bailey, T.B., Dinwiddie, D., Banks, O.G.B., McKnight, L.E., and McKnight, J.N., Basis of specificity for a conserved and promiscuous chromatin remodeling protein. *Elife*, *2021*. ***10****: p. e64061*.
76. Messenguy, F. and Dubois, E., Regulation of arginine metabolism in *Saccharomyces cerevisiae*: a network of specific and pleiotropic proteins in response to multiple environmental signals. *Food technology and biotechnology*, *2000*. ***38****(4): p. 277-286*.
77. Degols, G., Jauniaux, J.-C., and Wiame, J.-M., Molecular characterization of transposable-element-associated mutations that lead to constitutive L-ornithine aminotransferase expression in *Saccharomyces cerevisiae*. *European Journal of Biochemistry*, *1987*. ***165****(2): p. 289-296*.
78. Brody, L.C., Mitchell, G.A., Obie, C., Michaud, J., Steel, G., Fontaine, G., Robert, M.F., Sipila, I., Kaiser-Kupfer, M., and Valle, D., Ornithine delta-aminotransferase mutations in gyrate atrophy. Allelic heterogeneity and functional consequences. *Journal of Biological Chemistry*, *1992*. ***267****(5): p. 3302-3307*.
79. Messenguy, F. and Wiame, J.-M., The control of ornithinetranscarbamylase activity by arginase in *Saccharomyces cerevisiae*. *Febs Letters*, *1969*. ***3****(1): p. 47-49*.
80. Qian, W., Ma, D., Xiao, C., Wang, Z., and Zhang, J., The genomic landscape and evolutionary resolution of antagonistic pleiotropy in yeast. *Cell reports*, *2012*. ***2****(5): p. 1399-1410*.
81. Auesukaree, C., Damnernsawad, A., Kruatrachue, M., Pokethitiyook, P., Boonchird, C., Kaneko, Y., and Harashima, S., Genome-wide identification of genes involved in tolerance to various environmental stresses in *Saccharomyces cerevisiae*. *Journal of Applied Genetics*, *2009*. ***50****(3): p. 301-310*.
82. Fujita, K., Matsuyama, A., Kobayashi, Y., and Iwahashi, H., The genome-wide screening of yeast deletion mutants to identify the genes required for tolerance to ethanol and other alcohols. *FEMS yeast research*, *2006*. ***6****(5): p. 744-750*.
83. Tigano, M., Ruotolo, R., Dallabona, C., Fontanesi, F., Barrientos, A., Donnini, C., and Ottonello, S., Elongator-dependent modification of cytoplasmic tRNALysUUU is required for mitochondrial function under stress conditions. *Nucleic acids research*, *2015*. ***43****(17): p. 8368-8380*.
84. Furukawa, K., Innokentev, A., and Kanki, T., Regulatory Mechanisms of Mitochondrial Autophagy: Lessons From Yeast. *Frontiers in Plant Science*, *2019*. ***10***.
85. Backues, S.K., Lynch-Day, M.A., and Klionsky, D.J., The Ume6-Sin3-Rpd3 complex regulates *ATG8* transcription to control autophagosome size. *Autophagy*, *2012*. ***8****(12): p. 1835-1836*.
86. Gorter de Vries, A.R., de Groot, P.A., van den Broek, M., and Daran, J.-M.G., CRISPR-Cas9 mediated gene deletions in lager yeast *Saccharomyces pastorianus*. *Microbial cell factories*, *2017*. ***16****(1): p. 222*.
87. Gundelach, E., Suppressor studies on *ilv1* mutants of *Saccharomyces cerevisiae*. *Mutation Research/Fundamental and Molecular Mechanisms of Mutagenesis*, *1973*. ***20****(1): p. 25-33*.
88. Crabeel, M., Abadjieva, A., Hilven, P., Desimpelaere, J., and Soetens, O., Characterization of the *Saccharomyces cerevisiae ARG7* gene encoding ornithine acetyltransferase, an enzyme also endowed with acetylglutamate synthase activity. *European Journal of Biochemistry*, *1997*. ***250****(2): p. 232-241*.
89. Ernst, D.C. and Downs, D.M., Mmf1p couples amino acid metabolism to mitochondrial DNA maintenance in *Saccharomyces cerevisiae*. *Mbio*, *2018*. ***9****(1): p. e00084-18*.

CHAPTER 5

EXPLORATION OF mRNA-SIZED RNA IMPORT INTO *SACCHAROMYCES CEREVISIAE* MITOCHONDRIA BY A COMBINED SYNTHETIC BIOLOGY AND ADAPTIVE LABORATORY EVOLUTION APPROACH

Charlotte C. Koster, Kavish Kohabir, Maxime den Ridder, Marijke Luttik, Erik de Hulster, Martin Pabst, Pascale Daran-Lapujade

Abstract

Efficient gene integration using RNA-guided endonucleases has not yet been achieved in the mitochondrial genome. Import of nucleic acids into mitochondria, a controversial feature, is essential for implementation of Cas9-mediated genome engineering of mitochondria. Short RNA import naturally occurs in mitochondria, and several putative import mechanisms and determinants have been proposed. However to date, import of gene-length RNA, required for gene integration in the mitochondrial genome, has never been described. The goal of this study was to devise and test experimental strategies to detect and improve the import of mRNA-sized RNA in mitochondria, using *Saccharomyces cerevisiae* as model. A first fluorescence-based screening approach, relying on mitochondrial import of a fluorescent protein encoding mRNA revealed weak and stochastic RNA import, independent of the import signal fused to the mRNA. This screening also suggested a positive impact of mitochondrial co-import of the mRuby2 fluorescent protein with the mRNA. An adaptive laboratory evolution (ALE) approach, imposing a strong selection pressure for mRNA import to mitochondria, was then designed and tested to improve mitochondrial mRNA import. While the ALE approach did not improve mitochondrial mRNA import in the present study, it is a promising, unambiguous method for future studies testing different RNAs or mutants.

5.1 Introduction

The development of the CRISPR/Cas9 technology has revolutionized genome editing [1]. However, even the most genetically accessible eukaryotes harbor a genome that is not yet amenable for RNA-guided endonuclease (RGEN) editing: the mitochondrial genome. Mitochondria perform many essential cellular functions including the conservation of metabolic energy in the form of ATP, the maintenance of the cellular redox state, as well as production of co-factors, amino acids and lipids [2]. It is generally accepted that mitochondria have an endosymbiotic origin [3]. As a result, mitochondria have their own transcription and translation machineries, as well as a reduced genome encoding ribosomal subunits, a full set of tRNAs and several subunits of the respiratory chain. 1 % of the mitochondrial proteome is encoded on this mitochondrial genome, the rest is imported from the cytosol [4-6]. Apart from their key role in many eukaryotic cellular processes, mitochondria are also the source of many human diseases: over 250 pathogenic mutations in the mitochondrial DNA (mtDNA) have been identified [7]. Therefore, there have been substantial efforts in engineering mitochondrial genomes mostly focused on therapeutic applications [8, 9]. These efforts have led to the development of protein-based DNA editing strategies including base editors and zinc-finger- or TALE nucleases [10-13], which are limited to base exchange or depletion of mutated heteroplasmic mtDNA variants. More extensive modifications of mtDNA, such as gene integration or targeted deletion, would enable the reprogramming of mitochondrial genomes. This could either be applied to treat diseases, to endow mitochondria with new functions for metabolic engineering purposes, or even use mitochondria as chassis for the construction of synthetic cells. However, the large spectrum of applications of engineering of mitochondrial genomes is impeded by their poor genetic accessibility.

Currently, only microprojectile bombardment of DNA-coated metal particles (biolistic transformation) enables engineering of the mtDNA beyond base editing. This invasive method of mtDNA editing is restricted to two organisms, the yeast *Saccharomyces cerevisiae*, and the microalga *Chlamydomonas reinhardtii* [14, 15]. Expanding the mitochondrial genome editing 'toolbox' to include the flexible and less invasive CRISPR/RGEN editing has proven very challenging. This is largely explained by the three-component nature of CRISPR/RGEN systems. Unlike the existing mitochondrial programmable endonucleases, RGEN function depends on nucleic acids [16]. A guide RNA (gRNA) is required to direct the RGEN towards the editing site, and the induced double-stranded break can be repaired by a DNA molecule (repair DNA). While targeting proteins to the mitochondrial matrix is achieved by addition of mitochondrial targeting sequences (MTS [17]), targeting DNA and RNA molecules presents a substantial challenge. Mitochondria are not naturally competent for DNA, but native uptake of RNA in mitochondria has been reported for tRNAs in yeasts, plants and mammals (reviewed by Schneider [18]). Also, the ribonucleic subunits of RNases P and MRP, and

rRNA were reported to be imported in mammalian mitochondria [19-21]. The mechanistic understanding of mitochondrial RNA import is limited to the import mechanism of yeast $tRNA^{Lys}$(CUU). Here, the tRNA is first recruited to the mitochondrial membrane by binding Eno2p and subsequently bound by the precursor of mitochondrial lysyl-tRNA synthetase (pre-LysRS). Pre-LysRS is imported in mitochondria and likely co-imports the associated $tRNA^{Lys}$(CUU) [22]. This led to the hypothesis that certain RNA-sequences or structures of the RNA promote RNA import. Hence, multiple RNA 'import sequences' were derived from $tRNA^{Lys}$(CUU), 5S rRNA, mammalian RNAse P and RNase MRP that reportedly improve RNA import *in vitro* and occasionally *in vivo* [23-27].
Several studies have described successful gRNA targeting to the mitochondria through addition of RNA import sequences, reportedly leading to successful Cas9 editing of mtDNA [28-32]. In most instances, this was measured by monitoring mtDNA depletion, in lieu of a DNA repair. However, no definitive proof of RGEN-induced DNA-editing in the mitochondria, such as a site-specific shift in heteroplasmy, has been demonstrated. Since discrete and reproducible evidence of mitochondrial RNA localization and site-specific Cas9 activity on the mtDNA is lacking, several studies question the occurrence of gRNA import and mtDNA editing by Cas9 altogether [9, 10, 33-35]. These contrasting reports highlight the difficulty of the experimental design required to demonstrate if RNA is indeed imported into mitochondria and if site-specific RGEN-mediated mtDNA cleavage occurs. Even if gRNA import is successfully achieved, it is not sufficient for gene replacement or the addition of new functions in the mtDNA. This requires the integration of gene-length exogenous DNA, which is not natively imported in the mitochondria. However, repair DNA could be generated from mitochondrially-targeted RNA, through a mitochondrially-localized reverse transcriptase [36-38]. The use of such an RNA-mediated repair system would require the targeting of gene-length RNA to the mitochondria. There is little knowledge on the ability of mitochondria to import mRNA-sized RNA, as most studies focus on the import of short tRNA and gRNA (100 - 200 nucleotides), while the import of a 1000-nucleotide long mRNA was only reported once [25].

The goal of the present study is to devise and test experimental strategies to detect and enable the import of mRNA-sized RNA in mitochondria, using *S. cerevisiae* as paradigm. *S. cerevisiae* is often used as model organism in mitochondrial biology as it is one of the few model eukaryotes that can survive (extensive) mtDNA damage [39]. Additionally, *S. cerevisiae* is a robust, genetically tractable, and well-characterized model eukaryote, making it a perfect host to explore RNA import to mitochondria and mtDNA editing. RNA import in mitochondria was first explored by screening for mitochondrial localization of an mRNA fused to a library of RNA import signals and its fluorescent translation product. Next to this systematic approach, an adaptive laboratory evolution strategy was devised in an attempt to select for mutants with improved mitochondrial RNA import.

5.2 Materials & Methods

Strains, medium and cultivation

Saccharomyces cerevisiae strains used in this study were derived from a CEN.PK genetic background [40]. Unless indicated otherwise, yeast strains (Table 5.1) were grown aerobically at 30 °C in 500 mL shake flasks containing 100 mL minimal synthetic medium with ammonium as nitrogen source (SM) supplied with vitamins and trace elements, prepared and sterilized as described previously [37], in an Innova incubator shaker (Eppendorf AG, Hamburg, Germany) set at 200 rpm. Media were autoclaved at 121 °C for 20 minutes. Media for respiro-fermentative growth in shake-flasks contained 2 % (w/v) D-glucose, for respiring cultures 2 % (v/v) ethanol (SME) or 2 % (v/v) ethanol and 2 % (v/v) glycerol (SMEG) served as carbon sources. Glucose solutions (50 % w/v) and glycerol solutions (99 % w/v) were autoclaved separately for 10 minutes at 110 °C. Where relevant, media were complemented with Hygromycin B Gold (Invivogen, San Diego, CA, USA) to a final concentration 200 mg L^{-1}. Uracil- or arginine auxotrophic strains were supplemented with a separately sterilized solution of uracil (Ura) to a final concentration of 150 mg L^{-1} [41] or with L-arginine·HCl (Arg) to a final concentration of 450 mg L^{-1} of L-arginine [42]. For cloning, YPD medium, containing 10 g L^{-1} Bacto yeast extract, 20 g L^{-1} Bacto peptone and 20 g L^{-1} glucose was used, and when necessary supplemented with 200 mg L^{-1} G418 (Geneticin, Invivogen) or 200 mg L^{-1} hygromycin. For solid media 2 % (w/v) Bacto Agar

Table 5.1 Strains used in this study. The genotype of plasmids is shown between brackets, RNA-targeting signals are indicated in bold. * Indicates a mix of targeting signals.

Strain	*Relevant genotype*	*Strain background*	*Mitochondrially targeted protein*	*Mitochondrially targeted RNA (5'-3')*	*Source*
CEN.PK113-7D	*MATa URA3 LEU2 TRP1 HIS3*	-	-	-	[40]
CEN.PK113-5D	*MATa ura3-52 LEU2 TRP1 HIS3*	-	-	-	[40]
IMX2100	*MATa sga1::Spycas9-natNT2 Dura3*	CEN.PK113-7D	-	-	[42]
IMK1031	*ΔARG8 Δcar2 Δura3*	IMX2100	-	-	[42]
IMC111	*ura3-52 pUDC191(pCCW12-mRuby2-tENO1 URA3)*	CEN.PK113-5D	-	-	[43]
IMC112	*ura3-52 pUDC192(pTEF2-mTurquoise2-tSSA1 URA3)*	CEN.PK113-5D	-	-	[43]
IMC113	*ura3-52 pUDC193(pTEF1-Venus-tTDH1 URA3)*	CEN.PK113-5D	-	-	[43]
IMC159	*ura3-52 pUDC286(pTDH3-preCOX4-mRuby2-tADH1 URA3)*	CEN.PK113-5D	preCOX4-mRuby2	-	[44]
IMC173	*Δura3 pUDC329(pTEF1-preSU9-ymTq2-tENO2)*	IMX2100	preSU9-mTq2	-	[45]
IMC250	*ΔARG8 Δcar2 Δura3 + pUDC434(pTDH3-preCOX4MTS-ARG8-tADH1)*	IMK1031	Arg8	-	[42]
IMX2796	*ΔARG8 Δcar2 Δura3 yprctau3::pTDH3-preCOX4-mRuby2-tADH1*	IMK1031	preCOX4-mRuby2	-	This study
IME529	*pUDE970(pPGK1-HH-DF-mito-mTq2-HDV-tADH1, HygR)*	IMC159	preCOX4-mRuby2	*DF-mito-mTq2*	This study
IME530	*pUDE971(pPGK1-HH-D-mito-mTq2-HDV-tADH1, HygR)*	IMC159	preCOX4-mRuby2	*D-mito-mTq2*	This study
IME531	*pUDE972(pPGK1-HH-F-mito-mTq2-HDV-tADH1, HygR)*	IMC159	preCOX4-mRuby2	*F-mito-mTq2*	This study
IME532	*pUDE973(pPGK1-HH-h5S-mito-mTq2-HDV-tADH1, HygR)*	IMC159	preCOX4-mRuby2	*h5S-mito-mTq2*	This study
IME533	*pUDE974(pPGK1-HH-y5S-mito-mTq2-HDV-tADH1, HygR)*	IMC159	preCOX4-mRuby2	*y5S-mito-mTq2*	This study
IME534	*pUDE975(pPGK1-HH-hRP-mito-mTq2-HDV-tADH1, HygR)*	IMC159	preCOX4-mRuby2	*hRP-mito-mTq2*	This study
IME535	*pUDE976(pPGK1-HH-yRP-mito-mTq2-HDV-tADH1, HygR)*	IMC159	preCOX4-mRuby2	*yRP-mito-mTq2*	This study
IME536	*pUDE977(pPGK1-HH-hMRP-mito-mTq2-HDV-tADH1, HygR)*	IMC159	preCOX4-mRuby2	*hMRP-mito-mTq2*	This study
IME537	*pUDE978(pPGK1-HH-yMRP-mito-mTq2-HDV-tADH1, HygR)*	IMC159	preCOX4-mRuby2	*yMRP-mito-mTq2*	This study
IME538	*pUDE979(pPGK1-HH-mito-mTq2-D-HDV-tADH1, HygR)*	IMC159	preCOX4-mRuby2	*mito-mTq2-D*	This study
IME539	*pUDE980(pPGK1-HH-mito-mTq2-F-HDV-tADH1, HygR)*	IMC159	preCOX4-mRuby2	*mito-mTq2-F*	This study
IME540	*pUDE981(pPGK1-HH-mito-mTq2-h5S-HDV-tADH1, HygR)*	IMC159	preCOX4-mRuby2	*mito-mTq2-h5S*	This study
IME541	*pUDE982(pPGK1-HH-mito-mTq2-y5S-HDV-tADH1, HygR)*	IMC159	preCOX4-mRuby2	*mito-mTq2-y5S*	This study
IME542	*pUDE983(pPGK1-HH-mito-mTq2-hRP-HDV-tADH1, HygR)*	IMC159	preCOX4-mRuby2	*mito-mTq2-hRP*	This study
IME543	*pUDE984(pPGK1-HH-mito-mTq2-yRP-HDV-tADH1, HygR)*	IMC159	preCOX4-mRuby2	*mito-mTq2-yRP*	This study
IME544	*pUDE985(pPGK1-HH-mito-mTq2-hMRP-HDV-tADH1, HygR)*	IMC159	preCOX4-mRuby2	*mito-mTq2-hMRP*	This study
IME545	*pUDE986(pPGK1-HH-mito-mTq2-yMRP-HDV-tADH1, HygR)*	IMC159	preCOX4-mRuby2	*mito-mTq2-yMRP*	This study
IME546	*pUDE987(pPGK1-HH-mito-mTq2-tracr-HDV-tADH1, HygR)*	IMC159	preCOX4-mRuby2	*mito-mTq2-tracr*	This study
IME547	*pUDE969(pPGK1-HH-mito-mTq2-HDV-tADH1, HygR)*	IMC159	preCOX4-mRuby2	*mito-mTq2*	This study
IME551	*pUDE972(pPGK1-HH-F-mito-mTq2-HDV-tADH1, HygR)*	CEN.PK113-7D	-	*F-mito-mTq2*	This study
IME553	*pUDE974(pPGK1-HH-y5S-mito-mTq2-HDV-tADH1, HygR)*	CEN.PK113-7D	-	*y5S-mito-mTq2*	This study
IME556	*pUDE977(pPGK1-HH-hMRP-mito-mTq2-HDV-tADH1, HygR)*	CEN.PK113-7D	-	*hMRP-mito-mTq2*	This study
IME558	*pUDE979(pPGK1-HH-mito-mTq2-D-HDV-tADH1, HygR)*	CEN.PK113-7D	-	*mito-mTq2-D*	This study

IME567	*pUDE969(pPGK1-HH-mito-mTq2-HDV-tADH1, HygR)*	CEN.PK113-7D	-	*mito-mTq2*	This study
IME759	*pUDE1230(pTEF1-HH-DF-Cox2UTL-mitoARG8-dodecamer-HDV-tENO2, URA3)*	IMX2796	preCOX4-mRuby2	*DF-mito-ARG8*	This study
IME760	*pUDE1231(pTEF1-HH-D-Cox2UTL-mitoARG8-dodecamer-HDV-tENO2, URA3)*	IMX2796	preCOX4-mRuby2	*D-mito-ARG8*	This study
IME761	*pUDE1232(pTEF1-HH-F-Cox2UTL-mitoARG8-dodecamer-HDV-tENO2, URA3)*	IMX2796	preCOX4-mRuby2	*F-mito-ARG8*	This study
IME762	*pUDE1233(pTEF1-HH-h5S-Cox2UTL-mitoARG8-dodecamer-HDV-tENO2, URA3)*	IMX2796	preCOX4-mRuby2	*h5S-mito-ARG8*	This study
IME763	*pUDE1234(pTEF1-HH-y5S-Cox2UTL-mitoARG8-dodecamer-HDV-tENO2, URA3)*	IMX2796	preCOX4-mRuby2	*y5S-mito-ARG8*	This study
IME764	*pUDE1235(pTEF1-HH-hRP-Cox2UTL-mitoARG8-dodecamer-HDV-tENO2, URA3)*	IMX2796	preCOX4-mRuby2	*hRP-mito-ARG8*	This study
IME765	*pUDE1236(pTEF1-HH-yRP-Cox2UTL-mitoARG8-dodecamer-HDV-tENO2, URA3)*	IMX2796	preCOX4-mRuby2	*yRP-mito-ARG8*	This study
IME766	*pUDE1237(pTEF1-HH-hMRP-Cox2UTL-mitoARG8-dodecamer-HDV-tENO2, URA3)*	IMX2796	preCOX4-mRuby2	*hMRP-mito-ARG8*	This study
IME767	*pUDE1238(pTEF1-HH-yMRP-Cox2UTL-mitoARG8-dodecamer-HDV-tENO2, URA3)*	IMX2796	preCOX4-mRuby2	*yMRP-mito-ARG8*	This study
IME768	*pUDE1239(pTEF1-HH-F1-Cox2UTL-mitoARG8-dodecamer-HDV-tENO2, URA3)*	IMX2796	preCOX4-mRuby2	*F1-mito-ARG8*	This study
IME769	*pUDE1240(pTEF1-HH-tRK2tr93-Cox2UTL-mitoARG8-dodecamer-HDV-tENO2, URA3)*	IMX2796	preCOX4-mRuby2	*tRK2-mito-ARG8*	This study
IME770	*pUDE1244(pTEF1-HH-GFPdropout-Cox2UTL-mitoARG8-dodecamer-HDV-tENO2, URA3)*	IMX2796	preCOX4-mRuby2	*mito-ARG8*	This study
IME797	*pUDE971(pPGK1-HH-D-mito-mTq2-HDV-tADH1, HygR)*	IMC250	preCOX4-Arg8	*D-mito-mTq2*	This study
IME798	*pUDE972(pPGK1-HH-F-mito-mTq2-HDV-tADH1, HygR)*	IMC250	preCOX4-Arg8	*F-mito-mTq2*	This study
IME799	*pUDE974(pPGK1-HH-y5S-mito-mTq2-HDV-tADH1, HygR)*	IMC250	preCOX4-Arg8	*y5S-mito-mTq2*	This study
IME800	*pUDE977(pPGK1-HH-hMRP-mito-mTq2-HDV-tADH1, HygR)*	IMC250	preCOX4-Arg8	*hMRP-mito-mTq2*	This study
IME801	*pUDE979(pPGK1-HH-mito-mTq2-D-HDV-tADH1, HygR)*	IMC250	preCOX4-Arg8	*mito-mTq2-D*	This study
IME802	*pUDE969(pPGK1-HH-mito-mTq2-HDV-tADH1, HygR)*	IMC250	preCOX4-Arg8	*mito-mTq2*	This study
IMS1248	Chemostat replicate A population (562 generations)	IME759-IME770	preCOX4-mRuby2	**-mito-ARG8*	This study
IMS1249	Chemostat replicate B population (562 generations)	IME759-IME770	preCOX4-mRuby2	**-mito-ARG8*	This study
IMS1251	Chemostat replicate B single colony isolate 1 (350 generations)	IMS1249	preCOX4-mRuby2	**-mito-ARG8*	This study
IMS1252	Chemostat replicate B single colony isolate 2 (350 generations)	IMS1249	preCOX4-mRuby2	**-mito-ARG8*	This study
IMS1253	Chemostat replicate B single colony isolate 3 (350 generations)	IMS1249	preCOX4-mRuby2	**-mito-ARG8*	This study
IMS1254	Chemostat replicate B single colony isolate 4 (350 generations)	IMS1249	preCOX4-mRuby2	**-mito-ARG8*	This study
IMS1261	Chemostat replicate B population unevolved, sample taken after batch phase (“0 generations”)	IME759-IME770	preCOX4-mRuby2	**-mito-ARG8*	This study

(Becton Dickinson (BD), Franklin lakes, NJ, USA) was added to the medium prior to heat sterilization. Single-colony isolates were obtained by re-streaking single colonies three consecutive times on selective medium. Culture densities at any point of this study were determined using a Jenway™ 7200 spectrometer (Cole-Parmer, Staffordshire, UK) measuring the optical density at 660 nm (OD_{660}) with a 1 cm light path.

Escherichia coli XL1-Blue Subcloning Grade Competent Cells (Agilent Genomics, Santa Clara, US) were used for molecular cloning purposes and plasmid proliferation. Bacterial cultures were grown in Lysogeny Broth (LB) media (10 g L^{-1} Bacto™ Tryptone, 5 g L^{-1} Bacto™ Yeast Extract (BD), 5 g L^{-1} NaCl). LB media were autoclaved at 121 °C for 20 minutes. Where relevant, media were supplemented with ampicillin or chloramphenicol to a final concentration 10 mg L^{-1}. 5 mL liquid cultures were grown in a 15 mL CELLSTAR tube (Greiner Bio-One, Kremsmünster, Austria) at 37 °C while shaking 200 rpm in an Innova 4000 Incubator Shaker (Eppendorf).

Frozen stocks were prepared by addition of sterile glycerol (30 % v/v) to late exponential phase shake-flask cultures of *S. cerevisiae* or *E. coli* and 1 mL aliquots were stored aseptically at –80 °C.

General molecular biology techniques

Golden Gate assembly was done with 30 fmol of each fragment and 10 fmol of backbone. Additionally, the reaction mixtures contained 0.5 µL BsmbI or BsaI-HF v2 (New England Biolabs (NEB), Ipswich, MA, USA); 0.5 µL T7 DNA ligase (NEB); 1 µL T4 DNA ligase buffer (NEB) and was filled to 10 µL with nuclease-free water. When cloning with BsmbI or BsaI-HF v2, respective digestion temperatures of 42 °C and 37 °C were used. Ligation steps were always performed at 16 °C. The Golden Gate protocol used consisted of 25 cycles of 2 minutes digestion and 5 minutes ligation, followed by a final digestion step at 60 °C and a heat inactivation step at 80 °C, each lasting 10 minutes. Gibson isothermal assembly was performed using NEBuilder HiFi DNA Assembly Master Mix (NEB) with 10 fmol of backbone and 20 fmol of insert, and incubated at 50 °C for 1 hour. High fidelity PCR reactions were executed with Phusion High Fidelity Polymerase (Thermo Fisher Scientific) and PAGE-purified primers (Sigma-Aldrich, St. Louis, MO, USA) according to manufacturer's instructions. Gibson assembly products and Golden Gate assembly products were heat-shock transformed to chemical competent *E. coli* XL1-Blue cells, prepared following the supplier's instructions (Agilent Genomics). Plasmids were harvested from overnight bacterial cultures using GeneJET Plasmid Miniprep Kit (Thermo Fisher Scientific) according to the manufacturer's instructions. Fidelity of plasmids was assessed through diagnostic PCR reactions, done using DreamTaq Master Mix (Thermo Fisher Scientific) and desalted primers (Sigma-Aldrich, Table S5.2) according to manufacturer's instructions. Plasmids were transformed into yeast following the LiAc/SS carrier DNA/PEG method [46].

Plasmid and strain construction

Plasmids and strains expressing mito-mTq2 mRNA

Plasmids pUDE969 – pUDE987 expressing mito-mTq2 mRNA were cloned using a Golden-Gate assembly approach according to the Yeast Toolkit (YTK) design of Lee *et al.* [47]. An extensive cloning strategy including primer numbers is depicted in Figure S5.1. A list of plasmids and part plasmids is described in Table S5.1. First, all RNA targeting signals were ordered as complementary ssDNA primer pairs containing YTK-type 3a BsaI flanks for 5' signals or type YTK-type 4a BsaI flanks for 3' signals respectively, their sequences are listed in Table S5.2. *In vitro* annealing of the signals was done by heating a 1:1 ratio of complimentary ssDNA oligos to 100 °C for 5 minutes and letting it cool down on the bench at room temperature. Recoded mito-mTq2 (see Supplementary information for sequence) was ordered as a synthetic gene fragment at GeneArt (Thermo Fisher Scientific, Landsmeer, the Netherlands), and was amplified with primers containing YTK-type 3b BsaI flanks for 5'signals or YTK-type 3 BsaI flanks for 3' or no signals respectively (Table S5.2). All parts also contained BsmBI flanks which were used to clone the signals as part plasmids in backbone pYTK001 using Golden Gate cloning as described above, yielding part plasmids pGGKp265 - pGGKp291 ([47], Table S5.1). Backbone pUDE929 (*pPGK1-HH-dKanMX-HDV-tADH1*, *bla*, *ColE1*, *HygR*, *2µ)* was linearized by PCR-amplification with primers annealing to the HH and HDV with YTK-type 3 and 4(b) BsaI flanks, respectively (detailed construction strategy is shown in Figure S5.1). The different part plasmids with import signals

and/or mito-mTq2 (pGGKp265 - pGGKp291) were then assembled in the linearized backbone pUDE929 using Golden Gate assembly, yielding plasmids pUDE969 – pUDE987 (Table S5.1). For co-expression with mRuby2, plasmids were transformed in yeast strain IMC159 [44], yielding strains IME529 – IME547. For sole expression of mito-mTq2 mRNA, plasmids were transformed in CEN.PK113-5D, yielding strains IME551 – IME567. For co-expression with *preCOX4-ARG8*, the plasmids were transformed in strain IMC250 [42], yielding strains IME797 – IME802.

Plasmids and strains expressing mito-*ARG8* mRNA

The plasmids expressing mito-*ARG8* mRNA (pUDE1230 - pUDE1241) were constructed from the plasmid pUD1202 (*HH-GFPdropout-Cox2UTL-mitoARG8-dodecamer-HDV)*, which was ordered as a synthetic gene fragment at GeneArt (Thermo Fischer Scientific), and the fragment was PCR amplified with primers annealing to the HH/HDV sequences of the fragment. A backbone was PCR-amplified from plasmid pUDC329, using primers annealing to the promoter and terminator with homology flanks to the HH/HDV ribozyme sequences, respectively (Figure S5.1). The fragment and the backbone were cloned together by Gibson isothermal assembly, resulting in plasmid pUDC365. The full construct including promoter and terminator (*pTEF1-HH-GFPdropout-Cox2UTL-mitoARG8-dodecamer-HDV-tENO2*) was amplified and inserted via Gibson assembly into a PCR-amplified backbone of pUD518, which is a multi-copy vector, yielding plasmid pUDE1241 (Figure S5.1). The 5' signals were subsequently cloned into the GFP-dropout site of the plasmid as described by Lee *et al.* [47] to generate a library of mito-*ARG8* mRNA with different signals (plasmids pUDE1230 - pUDE1240, Figure S5.1, Table S5.1). Plasmids were transformed into yeast yielding strains IME759 – IME770 (Table 5.1).

Genomic integration

For genomic integration of the preCOX4-mRuby2 in the *YPRCtau3* locus in strain IMX2796, the preCOX4-mRuby expression cassette repair fragment was amplified from plasmid pUDC286 using high-fidelity PCR with primers that attached 60 bp homology flanks of the *YPRCtau3* locus of CEN.PK113-7D to the repair fragment (Table S5.1, Table S5.2). The repair fragment was co-transformed with plasmid pUDR514 expressing two gRNA sequences targeting the *YPRCtau3* locus in SpyCas9-expressing strain IMK1031 as described [48, 49], following the LiAc/SS carrier DNA/PEG method [46].

mRNA-FISH

Probe sets for single-molecule RNA FISH were designed using Stellaris Probe Designer (version 4.2; Biosearch™ Technologies Inc., Petaluma, CA). The sets targeting mito-mTq2 mRNA and *COX3* mRNA respectively consisted out of 25 Quasar-570 labelled probes and 24 CAL Fluor Red 635 labelled probes. All probes were 22-mers, designed to anneal with a minimal spacing of 1 nucleotide. Probes were diluted and stored as instructed by the manufacturer.

The protocol for RNA FISH was adopted from Schwabe and Bruggeman [50] with the following modifications. 20 mL starter cultures (SMEG + HygB) were inoculated and after 6-8 hours and transferred to 100 mL fresh medium. Approximately $1{\cdot}10^9$ cells were sampled from exponentially growing cultures and fixed in 4 % (w/v) paraformaldehyde in SM. Formaldehyde fixation was done while shaking 200 rpm at 30 °C for at least 30 minutes, but for no longer than 1 hour. Fixed cells were stored at least 1 night at 4 °C. All consecutive steps entailed RNase-free work, for which a separate bench was dedicated for work with ambidextrous glove protection. Where possible, chemicals and consumables were autoclaved at 121 °C for at least 45 minutes or otherwise treated with RNase*Zap* decontamination solution (Sigma-Aldrich). Prior to spheroplasting, cells were washed twice with cooled spheroplasting buffer (1.2 M sorbitol, 0.1 M pH 7.5 potassium phosphate buffer (KPB)). Cells were resuspended in 1.25 mL spheroplasting buffer containing approximately 800 U lyticase from *Arthrobacter luteus* (L4025-50KU; Sigma-Aldrich) and incubated at 30 °C while shaking gently. Progression of spheroplast formation was monitored by phase-contrast microscopy, aiming for roughly 50 % of the cells to appear phase-contrast dark (i.e. after 50 – 70 min). At any succeeding point of the protocol, centrifugation of spheroplasts was done at 380 × *g* for 5 minutes at room temperature and resuspension of spheroplasts was done by gently flicking the tube. The spheroplasting reaction was terminated by 2 cycles of spinning down the spheroplasts and washing with 1 mL cold spheroplasting buffer. Obtained 1 mL

suspensions were distributed in 200 µL aliquots and pelleted. Spheroplasts were resuspended in 200 µL 70 % EtOH for storage at 4 °C until further use. Prior to hybridization, spheroplasts were spun down, ethanol supernatants were aspirated, and pellets were resuspended in 1 mL room-temperature (RT) wash buffer (2x saline sodium citrate (SSC) buffer, 10 % (v/v) formamide). After a 5-minute incubation on bench, samples were centrifuged, aspirated, and resuspended in 50 µL hybridization solution (100 g L^{-1} dextran sulfate sodium salt from *Leuconostoc* spp., 1 g L^{-1} *E. coli* tRNA, 2 mM vanadyl ribonucleoside complex (VRC), 200 mg L^{-1} bovine serum albumin (BSA), 2x SSC buffer, 10 % formamide, 125 nM probe) at RT. Hybridization was done overnight, protected from light, in a heat block set at 30 °C and 300 rpm. After hybridization, spheroplasts were washed with 1 mL wash buffer at RT, centrifuged and aspirated. Subsequently, pellets were resuspended in 1 mL wash buffer (RT) containing 500 ng mL^{-1} DAPI (Sigma-Aldrich) and incubated for 30 minutes at 30 °C, after which the samples were pelleted and aspirated. Spheroplasts were resuspended in 1 mL wash buffer and incubated for 30 minutes at 30 °C.
Spheroplasts were resuspended in 100 µL gelvatol (155 g L^{-1} polyvinyl alcohol (PVA), 21 % (v/v) glycerol, 53 % (v/v) Tris buffer (0.2 M Trizma base (Sigma Aldrich), pH 8.5), a few crystals of sodium azide) as a mounting agent. 5 µL of this suspension was placed on a SuperFrost objective slide (Menzel-Gläzer, Thermo Fisher Scientific) and covered with a #1.5 high precision cover slip (Paul Marienfeld & Co., Lauda-Königshofen, Germany). Coverslips were sealed to objective slides using transparent nail polish (product no. 11244545; HEMA, Amsterdam, the Netherlands). Slides were labelled and stored at room temperature while protected from light until required for microscopy.

Microscopy

Brightfield, phase-contrast and fluorescence widefield microscopy were done using Zeiss Axio Imager Z1 Upright Microscope (Carl Zeiss, Oberkochen, Germany), equipped with a HAL 100 Halogen Illuminator (Carl Zeiss) and an HBO 100 illuminating system (Carl Zeiss). Acquisitions were done with an AxioCam HRm Rev3 detector (60N-C 1'' 1.0x) (Carl Zeiss), connected to Zen 2 v10.0.0.910 (Carl Zeiss) software. 1 mL from a liquid culture was centrifuged for 3 minutes 3000 × *g* at RT, and resuspended in sterile dH_2O, of which 4 µL was imaged using a 100x EC Plan-NeoFLUAR lateral magnification oil-immersion objective with a numerical aperture (NA) of 1.3 (Carl Zeiss) after oil immersion with Immersol™ 518F type F immersion oil (Carl Zeiss). mTurquoise fluorescence was detected using filter set 47 (Carl Zeiss AG; excitation bandpass (BP) filter 436/20 nm, beam splitter filter 455 nm, emission filter BP 480/40 nm, Figure S5.2). The exposure time was always set at 1200 ms for all strains expressing mito-mTq2 mRNA. mRuby2 fluorescence was detected using filter set 14 (Carl Zeiss, excitation BP 535/25, emission Long Pass (LP) 590, Figure S5.2), with an exposure time of 500 ms. Exposure times were set such that the red fluorescence was clearly visible, and no blue fluorescence (caused by e.g., auto-fluorescence or overexposure of the cells) was observed.
For image acquisition of prepared slides as described for FISH experiments, fluorescence widefield microscopy was performed using an Olympus IX 81 inverted microscope (Olympus, Tokyo, Japan), equipped with an Andor™ AMH-200-F6S high-power metal halide lamp (Oxford Instruments, Abingdon, UK) and an Andor™ Luca R EM-CCD camera (Oxford Instruments) connected to Andor™ iQ3 control and acquisition software (Oxford Instruments). Acquisitions were done using a 100x UplanFLN lateral magnification objective with an NA of 1.3 after oil immersion with IMMOIL-F30CC low auto-fluorescence type F immersion oil (Olympus).

Image analysis

Image analysis was performed using the FIJI package of ImageJ [51]. Mean co-localized fluorescence of mRuby and mTurquoise2 in the mitochondria was analysed by thresholding the red fluorescence channel of each image using the minimal algorithm to determine the location of the mitochondria in the image. Based on the thresholded red fluorescence image, a binary mask was generated indicating the regions of interest (ROI, i.e. mitochondrial structures) of the image. Subsequently, the mask was overlaid on the red and blue fluorescent channels of the image and the mean fluorescence for each ROI was determined using FIJI, resulting in a mean mitochondrial fluorescence per cell in both the red and blue channel. Correlation coefficients and ANOVA one-way analysis were calculated using R-based analysis in RStudio (RStudio, Boston, MA, USA) using the ggpubr package of R.
When there was no fluorescent protein targeted to the mitochondrial matrix, masking was performed on the brightfield channel of the microscopy data. The cell boundary was determined in the brightfield image

using auto-thresholding with the Otsu algorithm. The background value was determined by measuring the fluorescence in the non-masked area (i.e. without cells) Subsequently, within the cell area, the background value was subtracted and the mean fluorescence in the whole cell was determined. For strains expressing mito-mTq2 mRNA in combination with preCOX4-Arg8p, a fluorescence threshold was set at 1500, as this was found to be the value above which auto-fluorescence could not be detected in the negative controls.

FISH images were subject to deconvolution. The generation of Point Spread Functions (PSFs) for spectral deconvolution was done in ImageJ, using the Parallel Spectral Deconvolution (v1.9) plugin. All settings were kept constant, except for stack height in number of slices, spacing between slices in nm and emission wavelength (λ_{EM}) of each registered channel. For each of these variations, a new PSF needed to be generated to subsequently deconvolve the image accordingly. All images were normalized by setting the sum of pixels to 1. Algorithmic deconvolution was done in ImageJ using the Deconvolutionlab2 (v2.1.2) plugin. Each channel was deconvolved in grey-scale separately and finally combined to a multi-channel, multidimensional acquisition. Deconvolution was done using Richardson-Lucy deconvolution, commonly used for deconvolution when the PSF is known, but no or little information is known about the noise. All deconvolutions for this study were done with 35 iterations.

Growth characterization of mito-*ARG8* mRNA mutants

Growth rate analysis of mutants expressing mito-*ARG8* mRNA was performed in 96-wells microtiter plates at 30 °C and 250 rpm using a Growth Profiler 960 (EnzyScreen BV, Heemstede, The Netherlands), essentially as described in Chapter 4 [42]. Frozen glycerol stocks were inoculated in 100 mL SMD + Arg and grown overnight. Cultures were supplied with uracil when necessary. 0.5 mL of the overnight culture was transferred to 100 mL SMD + Arg and grown until the OD_{660} had doubled at least once to ensure exponential growth. The cultures were spun down and washed twice in sterile dH_2O, then diluted to an OD_{660} of 15. A 96-wells microtiter plate (EnzyScreen, type CR1496dl) containing SMD or SMEG with final working volumes of 250 µL was inoculated with a starting OD_{660} of 0.3, approximating 300 cells. Growth rate analysis and the calculation of OD equivalents was performed as described by Boonekamp *et al.* [52].

Laboratory evolution and bioreactor cultivation

Chemostat cultures

Pre-cultures of IME759 – IME770 strains expressing mito-*ARG8* mRNA and preCOX4-mRuby2 were grown overnight in 20 mL volume of SMD + Arg in 100 mL shake flasks. A total of 20 OD_{660} units of each strain was combined in a single 50 mL centrifuge tube so an equal number of cells per strains would be inoculated in each bioreactor. The combined cultures were inoculated for automated evolutionary engineering in duplicate parallel 2 L Applikon Bio laboratory bioreactors (Getinge, Delft, the Netherlands) with a 1 L working volume [53], containing arginine-limited SME (SM supplemented with 20 g L^{-1} ethanol, 0.3 g L^{-1} antifoam Pluronic PE 6100 (BASF, Ludwigshafen, Germany) and 60 mg L^{-1} L-arginine). Cultures were stirred at 800 RPM, the temperature was controlled at 30 °C, and the pH was maintained at 5.0 through automated addition of 2.0 M KOH. The cultures were sparged with air at a flow rate of 475 mL min^{-1}. The strains were grown in batch until a stabilization of CO_2 indicated arginine depletion. Upon arginine depletion, evolution in continuous culture set-up was started by switching on medium pumps to obtain a constant in-flow of arginine-limited SME. The culture volume was kept constant at 1 L using an effluent pump that was controlled by an electric level sensor, resulting in a stable dilution rate of 0.1 h^{-1}. Evaporation of water and volatile metabolites was minimized by cooling the outlet gas of bioreactors to 4 °C in a condenser. CO_2 and O_2 concentrations were measured continuously in in the reactor outlet gas which was dried with a PermaPure PD-50T-12MPP dryer (Permapure, Lakewood, NJ) prior to analysis, and concentrations in the outlet gas were measured with an NGA 2000 Rosemount gas analyzer (Emerson, St. Louis, MO) calibrated with reference gas containing 3.03 % CO_2 and N6-grade N_2 (Linde Gas Benelux). All process parameters and output were continuously monitored *in situ* using Lucullus software (v3.8.2, Getinge) and through regular sampling and off-line measurements described below. Two independent bioreactors named A and B, corresponding to two evolution lines, were run in parallel. Glycerol stocks of the chemostat populations were taken at regular intervals at the same time as off-line measurements. The samples used for analysis were taken at the start of the chemostat phase

right after the batch phase ended (IMS1261) and at the very end of the chemostat cultivation prior to ending the experiment (562 generations, IMS1248 and IMS1249). Single-colony isolates IMS1250 - IMS1254 were isolated from samples taken from bioreactor B after approximately 350 generations. To obtain the isolates, a sample of bioreactor B was plated on SME without arginine, and four colonies were picked and restreaked on medium without arginine three consecutive times to obtain single-colony isolates.

Sequencing batch reactors

After 400 generations, 100 mL of bioreactor replicate B were inoculated in a sequencing batch reactor setup (SBR), consisting of a 2 L Applikon Bio laboratory bioreactor with a 1 L working volume. Aeration, pH, temperature, and dissolved oxygen concentration thresholds were the same as in the chemostat cultivations, but arginine was omitted in the SME medium. Optical density was measured in the bioreactor using a BE2100 non-invasive optical biomass sensor (Buglab LLC, Concord, CA, USA). The SBR culture was grown in batch mode until OD and CO_2 stabilized, upon which approximately 95 % of the culture volume was removed and the bioreactors were refilled with fresh medium [54]. This process was repeated for six batch cultivation cycles, after which it was stopped since no improvement was observed.

Characterization of IMS1254 physiology in batch culture

Single-colony isolate IMS1254 was characterized in batch culture with a setup similar to the SBR. The strain was grown overnight in a shake flask containing 100 mL SMD + Arg, washed with sterile dH_2O and inoculated in three replicate bioreactors run in parallel, one containing SME, and two containing SM supplemented with 20 g L^{-1} glucose. The cultures were grown in batch mode to deplete any leftover arginine in the medium, upon which am empty-refill cycle with the respective medium (SMD or SME, omitting arginine) was started manually. The cultures were then characterized in batch mode by online monitoring and off-line measurements.

Analytical methods for cultures in bioreactors

The bioreactors were sampled at regular intervals and several parameters were measured off-line. Biomass dry weight measurements of the bioreactor experiments were performed as previously described, using 10 mL of culture volume [55]. Metabolite concentrations in culture supernatants, obtained by centrifugation of 1 mL of culture for 3 min at 13.000 × *g*, were analyzed by high-performance liquid chromatography (HPLC) on an Agilent 1260 HPLC system (Agilent Technologies, Santa Clara, CA) fitted with a Aminex HPX 87H column (Bio-Rad, Hercules, CA) operated at 60 °C with 5 mM H_2SO_4 as the mobile phase with a flow rate of 0.6 mL min^{-1} and detection by refractive index and wavelength absorbance detectors at 214 nm. OD_{660} was determined on culture samples as described above. Colony forming units (CFU) were determined by making a 0 ×, 100 ×, 1000 × and 10.000 × serial dilutions of biomass samples in sterile dH_2O and plating on solid SMD, SMD + Arg, SME and SME + Arg and counting CFU manually. Cell counts were determined with a Z2 Coulter Counter (Beckman Coulter, Woerden, the Netherlands) using a 50 µm aperture. Particle volume was calibrated using 5 µm latex beads (Beckman Coulter) as recommended by the supplier. Appropriate dilutions were measured to have a coincidence between 5 and 10 %.

Whole genome sequencing (WGS)

To obtain DNA for WGS, glycerol stocks from evolved strains IMS1251-1254 were grown overnight in 100 mL SMD + Arg to late exponential phase (OD_{660} > 10) and harvested by centrifugation. Biomass of evolved bioreactor populations IMS1248 and IMS1249 was harvested directly from the outflow of the bioreactors and spun down. Genomic DNA was extracted with the Qiagen Blood & Cell Culture DNA kit (Qiagen, Germantown, MD), following manufacturer's specifications. WGS was performed by Macrogen Europe (Amsterdam, the Netherlands) on a Novaseq 6000 sequencer (Illumina, San Diego, CA) to obtain 151 cycle paired-end libraries with an insert-size of 550 bp using TruSeq Nano DNA library preparation, yielding 2 gigabases in total per sample. All Illumina sequencing data are available at NCBI (https://www.ncbi.nlm.nih.gov/) under the Bioproject accession number PRJNA923502. Reads were mapped using BWA (version 0.7.15) [56] to a IMX2600 reference [57] Alignments were processed using SAMtools (version 1.3.1) [58],

and variants were called by applying Pilon (version 1.18) [59]. Sequencing output was visualized using the Integrative Genomics Viewer [60] and chromosome copy numbers were determined using Magnolya [61].

Mitochondria isolation

Mitochondria required for Western blotting or proteomics were isolated from shake-flask cultures. Glycerol stocks were grown overnight on SM medium supplemented with glucose as a carbon source, supplemented with HygB for strains expressing mito-mTq2 mRNA and supplemented with arginine for strains obtained from bioreactor experiments. The pre-cultures were transferred to fresh medium to an OD_{660} of 0.1 in a total of 300 mL of SMEG, supplied with arginine or HygB when required. The strains were then grown to mid-exponential phase (OD_{660} of 4-8) and 300 mL of culture were harvested by centrifugation. Mitochondrial isolation was performed as described in Koster *et al.* [45, Chapter 3]. Mitochondria isolation was performed in biological duplicate experiments for each strain. The mitochondria-enriched fraction was immediately processed for Western blot analysis. For proteome analysis, the mitochondria were resuspended in 5 mL ice-cold Sorbitol-HEPES buffer (0.6 M sorbitol, 20 mM HEPES-KOH, 2 mM $MgCl_2$, cOmplete Protease Inhibitors Cocktail (Roche Diagnostics, Rotkreuz, Switzerland)), and distributed in 500 µL aliquots, equating approximately 1 - 2 mg total protein. The aliquots were flash-frozen in liquid nitrogen and stored at -80 °C for proteome analysis.

Western blotting

For protein expression analysis by western blot, 50 µL of isolated mitochondria were resuspended in 200 µL ice-cold lysis buffer (50 mM HEPES pH 7.5, 150 mM NaCl, 2.5 mM EDTA, 1 % v/v Triton X-100, cOmplete Mini Protease Inhibitors (Roche Diagnostics, Rotkreuz, Switzerland)), and thoroughly vortexed to release proteins. The supernatant was cleared by centrifugation (10.000 $\times$ *g* for 10 min at 4 °C) and protein concentrations were determined using the Quickstart Bradford Protein Assay (Bio-rad Laboratories, Inc, Hercules, CA, USA). 25 µg of total protein were loaded on two separate 4-15 % Mini-Protean TGX Stain-Free Protein gels (Bio-rad) and were run at 250 V for 20 min. Gels were activated using the stain free protocol on a ChemiDoc MP Imager (Bio-rad) to verify separation of the protein extracts. Proteins were transferred to PVDF membranes using a Turboblotter system (Bio-rad). Successful transfer of proteins was verified using the ChemiDoc MP Imager. Membranes were blocked for 1 hour in TBS + 1 % casein (Blocking Buffer, Bio-rad) and washed once with PBSt (phosphate saline buffer pH 7.2 with 0.05 % Tween20). Washed membranes were incubated overnight, shaken at 4 °C, in PBSt. For detection of subunit 3 of cytochrome *c* oxidase, anti-Cox3p antibody (from mouse, #459300, Invitrogen, Waltham, MA, USA) was added to one of the membranes of each replicate, diluted 1 in 1000 in PBSt. After overnight incubation, membranes were washed in PBSt. Membranes with anti-Cox3p were incubated with the secondary anti-mouse HRP antibody, diluted 1 in 2000 (from goat, Dako Agilent, Santa Clara, CA, USA) for 3 hours at 4 °C. The other membranes were incubated for detection of mTurquoise2 with anti-GFP Alexa Fluor 488 conjugated antibody (from rabbit, #A21311, Invitrogen) for 1 h at room temperature. HRP antibody levels were detected with enhanced chemiluminescence using the Clarity Max Western ECL Substrate (Bio-rad) and chemiluminescence- and fluorescence signals were detected using a ChemiDoc MP Imager (Bio-Rad).

Proteomic analysis

Proteome extraction and analysis was performed essentially as described in den Ridder *et al.* [62], with modifications to accommodate proteome extraction of isolated mitochondria. Two biological replicates were used for each experiment. Mitochondrial aliquots (500 µL, -80 °C) containing 1 - 2 mg protein were thawed on ice and resuspended in lysis buffer composed of 100 mM Triethylammonium bicarbonate (TEAB) containing 1 % SDS and phosphatase/protease inhibitors. Mitochondrial fractions were lysed by adding glass beads to the mitochondria and vortexing in 3 cycles of 1 minute, alternated with 1 minute rest on ice. Proteins were reduced by addition of 5 mM DTT and incubation for 1 hour at 37 °C. Subsequently, the proteins were alkylated for 60 minutes at room temperature in the dark by addition of 50 mM iodoacetamide. Protein precipitation was performed by addition of four volumes of ice-cold acetone (-20 °C), followed by 1 hour freezing at -20 °C. The proteins were solubilized using 100 mM ammonium bicarbonate. Proteolytic digestion was performed by Trypsin (Promega, Madison, WI), 1:100 enzyme to protein ratio, and incubated at 37 °C

overnight. Solid phase extraction was performed with an Oasis HLB 96-well µElution plate (Waters, Milford, USA) to desalt the mixture. Eluates were dried using a SpeedVac vacuum concentrator at 50 °C and frozen at -80 °C. An aliquot corresponding to approximately 1 µg protein digest was analyzed using a one-dimensional shot-gun proteomics approach in three technical replicate experiment for each biological replicate [62]. Data were analyzed against the proteome database from *Saccharomyces cerevisiae* (UniProt, strain ATCC 204508/ S288C, Tax ID: 559292, July 2020) using PEAKS Studio X (Bioinformatics Solutions Inc., Waterloo, Canada) as described [62]. The significance score for evaluating the observed abundance changes was calculated using a one-way ANOVA and expressed as the $-10 \cdot log_{10}(p)$, where p is the significance testing p-value, which represents the likelihood that the observed change is caused by random chance. Based on the average of the duplicate quantified proteins, the fold change of each protein in a specific condition was calculated relative to the unevolved sample. The average fold changes of the technical replicates were subsequently used to determine the standard deviations of the biological replicates. Results from the analysis are attached in SI file 2.

GO term enrichment

A functional term enrichment analysis was performed to determine whether specific Gene Ontology (GO [63]) terms or were shared between proteins that were significantly more- or less abundant in two out of three or all three strains. GO-term analysis was performed using the GO::TermFinder [64] accessed through the GO Term Finder webpage hosted by the *Saccharomyces* genome data base (SGD, https://www.yeastgenome.org, [65]) using GO version 2023-01-01. A list of genes of interest (i.e. mitochondrial proteins more abundant in all strains), was analyzed against a custom reference list containing all *S. cerevisiae* mitochondrial proteins (SI file 2). The reference list of mitochondrial proteins was obtained by exporting a list of all genes with "mitochondrion" as cellular component (GO:0005739) from SGD. The enrichment strength was calculated by dividing the number of proteins in the entered dataset that associated with a GO-term by the number of proteins associated with the same GO-term in the reference list.

5.3 Results

Fluorescence-based screening of RNA import in yeast mitochondria

An array of RNA structures of 20 – 80 nucleotides that aid import of RNA in mitochondria have been proposed [24, 28, 66]. A subset of these RNA structures was selected to test their ability to import longer RNA molecules (Table 5.2). Four tRNA-derived import signals (DF, D-arm, F-stem, F1-stem) identified in *S. cerevisiae* were selected based on their described ability to effectively import RNA in yeast mitochondria. Additionally, three signals described to aid import in human mitochondria (5S rRNA, RNase P and RNase MRP) were tested, as well as their three yeast homologues (Table 5.2, Figure S5.3). Lastly, since tracrRNA was also hypothesized to aid RNA import, this RNA structure was also tested, resulting in a library of eleven import signals, that could be cloned at either the 5' end or the 3' end of an mRNA.

Table 5.2. RNA targeting signals used in this study.

RNA signal (5' and/or 3')	*Length (nt)*	*Origin*	*Source*
DF	39	Yeast tRNALys *tRK1*	[23]
D-arm	16	Yeast tRNALys *tRK1*	[23]
F-stem	20	Yeast tRNALys *tRK1*	[23]
F1	25	Yeast tRNALys *tRK1*	[23]
h5S	46	Human 5S rRNA MAM domain	[26]
y5s[1]	14	Yeast 5S rRNA	This study
hRP	20	Human RNase P RNA component	[24]
yRP[1]	20	Yeast RNase P	This study
hMRP	20	Human Rnase *MRP* RNA component	[24]
yMRP[1]	20	Yeast RNase P	This study
tracrRNA[2]	79	CRISPR-Cas9 tracrRNA structure	[28]
tRK2-tr93[2]	76	Yeast tRNALys *tRK2*	[27]
No signal	0		

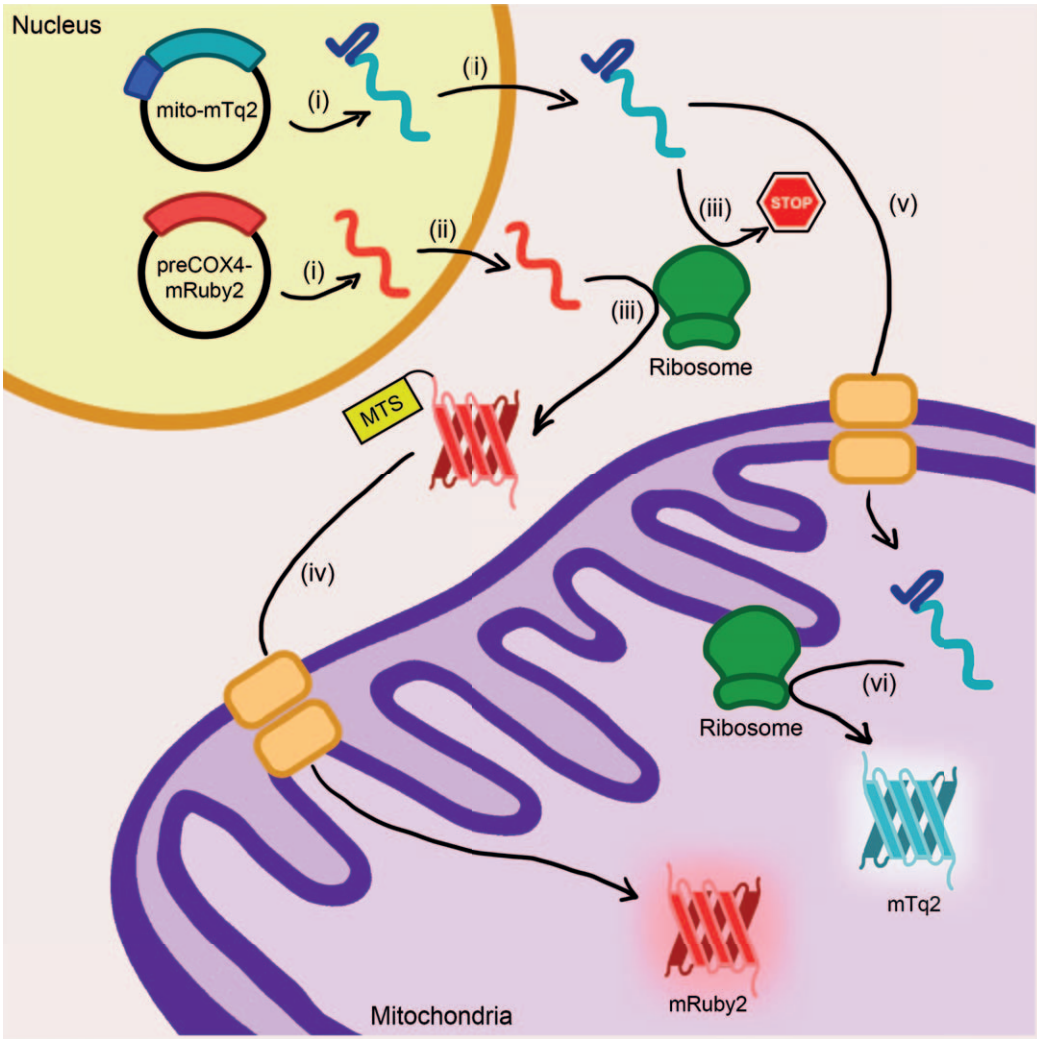


Figure 5.1. Schematic overview of the screening method to detect RNA import in mitochondria. (i) Mitochondrially recoded mTurquoise2 (mito-mTq2) fused to an RNA targeting signal (dark blue) is expressed from a plasmid, as well as mRuby (standard codon usage) fused to mitochondrial targeting signal preCOX4. (ii) Both mRNAs are expressed using Polymerase II promoters and therefore contain a 5' cap and poly(A) tail, ensuring export from the nucleus, which are cleaved off by ribozymes attached to the 5' and 3' end (not pictured). (iii) mRuby is translated by cytosolic ribosomes, the mito-mTq2 mRNA is mitochondrially recoded and therefore contains internal STOP codons, yielding a truncated protein when expressed in the cytosol. (iv) mRuby2 is fused to a mitochondrial targeting signal (MTS) preCOX4, leading to translocation of the protein upon translation to the mitochondrial matrix, which results in localized red fluorescence in the mitochondria. (v, vi) mito-mTq2 mRNA that is imported into mitochondria can be translated into the full mTq2 protein by the mitochondrial ribosomes, resulting in localized blue fluorescence in the mitochondria. Co-localization of both fluorescent signals indicates successful mitochondrial RNA import.

The RNA import capacity of eleven RNA structures was investigated by a fluorescence-based screening approach (Figure 5.1), relying on the alternative codon usage of yeast mitochondria. The UGA codon encodes a tryptophan (Trp) in mitochondria, while it encodes a STOP codon in cytosolic translation [67]. The 720-nucleotide mRNA sequence encoding the blue fluorescent protein mTurquoise2 (mTq2) was the chosen RNA molecule for mitochondrial import. The mTq2 sequence was recoded for mitochondrial translation by replacing the two native Trp-codons with the UGA codon. Translation of the recoded mRNA (referred to as mito-mTq2 mRNA) in mitochondria would result in a 238 amino acids fluorescent protein, while cytosolic translation would yield truncated, non-fluorescent proteins of 65 and 161 amino acids (Figure S5.4). The

selected RNA import signals were cloned at either the 5' or 3' end of the recoded mito-mTq2 mRNA in a multicopy plasmid under control of a strong constitutive promoter. This resulted in eighteen different strains, each expressing mito-mTq2 with a putative RNA import signal attached at either the 5' or the 3' end. A control omitting an import signal was included, leading to a total of nineteen plasmids (pUDE969-987, Table S5.1). To prevent potential interference with mitochondrial translation, the 5'-cap and poly(A)-tail added during transcription in the nucleus were removed by framing the 5' and 3' of the RNA construct with self-cleaving ribozymes (Figure S5.1) [68, 69]. Constitutive fluorescent labelling of mitochondria was achieved by targeting the cytosolically translated red fluorescent protein mRuby2 to mitochondria (preCOX4-mRuby2, Figure 5.1). The nineteen mito-mTq2 mRNA plasmids were expressed in the presence (strains IME529 – IME546) or absence of preCOX4-mRuby2 (strains IME551 – IME567), and localization of the red and blue fluorescent proteins was assessed by fluorescence microscopy.
Two negative controls, the parental strain CEN.PK113-5D expressing no fluorescent protein and IMC159 only expressing preCOX4-mRuby2, and the positive control IMC173 expressing the mTq2 protein targeted to the mitochondria with a preSU9-MTS (called preSU9-mTq2), but no mRuby2 protein, were included. As expected, the negative controls IMC159 and CEN.PK123-5D did not show any visible fluorescence in mitochondria in the blue channel, while the positive control IMC173 did, and only IMC159 showed red fluorescence, localized to the mitochondria (Figure 5.2A).

All strains co-expressing preCOX4-mRuby2 and the mito-mTq2 mRNA showed clear red fluorescence localized to the mitochondria. Additionally, blue fluorescence was observed with varying intensity, but was always co-localized with red fluorescence (Figure 5.2A, Figure S5.5), suggesting import and translation of the mito-mTq2 mRNA in the mitochondria. Blue fluorescence was in most strains not homogeneously distributed across all imaged cells. One strain, IME531, carrying a 5'-F stem RNA structure linked to the mito-mTq2 mRNA, consistently showed blue fluorescence in all imaged mitochondria, with strong overlap with red fluorescence, suggesting mito-mTq2 mRNA import into mitochondria. Surprisingly, IME547 in which the import signal was omitted, also showed increased fluorescent signals in the blue channel that co-localized with the red fluorescent mitochondria, thereby questioning the requirement of import signal for mitochondrial RNA import. Quantification of the fluorescent signals showed that, for most strains, mean blue fluorescence was not higher than in the negative control strains IMC159 and CEN.PK113-5D (Figure 5.2B, Figure S5.6), with the exception of IME531 (F-stem) and IME547 (No RNA import signal). For some strains, a fraction of the imaged cells did show substantial blue fluorescence (e.g., IME536 (hMRP, Figure 5.2B), IME537 and IME541 (5'-yMRP and 3'-y5S, resp., Figure S5.6), suggesting that mito-mTq2 mRNA import and translation in mitochondria might occur, but as a stochastic process.

Is RNA import to mitochondria correlated to the co-import of the mRuby2 protein?

Surprisingly, the blue fluorescent signal in strains solely expressing the mito-mTq2 mRNA was completely absent (Figure 5.2, Figure S5.5, Figure S5.6). In these strains, devoid of preCOX4-mRuby2, any signal measured above background levels (e.g., IME558) was clearly caused by background fluorescence, and no localization of fluorescence to mitochondria was observed (Figure S5.6). Conversely, all strains co-expressing the mito-mTq2 mRNA together with the preCOX4-mRuby2 displayed a high and significant Pearson correlation coefficient (PCC) between red and blue fluorescence above 0.5 (p-value < 0.05, Figure 5.2C), among which six strains had a PCC above 0.9. The positive control strain IMC173 for mTq2 did not show a significant PCC between red and blue channels, and the negative control expressing solely mRuby2 in the mitochondria showed a weak PCC below 0.5 ($p < 0.05$). The observed correlation between mTurquoise2 fluorescence and mRuby2 fluorescence could therefore not result from spectral overlap between the proteins. Further attempts to assess mito-mTq2 mRNA localization by RNA-FISH using mito-mTq2-targeting probes confirmed its expression and nuclear export, however its cellular localization was too scattered to draw conclusions on its presence in mitochondria (Figure S5.7).

mRNA localization to mitochondria proved technically challenging to unambiguously demonstrate by fluorescence microscopy and RNA-FISH. Therefore, the translation of mitochondrial and cytosolic alternative mito-mTq2 translation products was monitored through Western blotting. GFP has 95 % amino

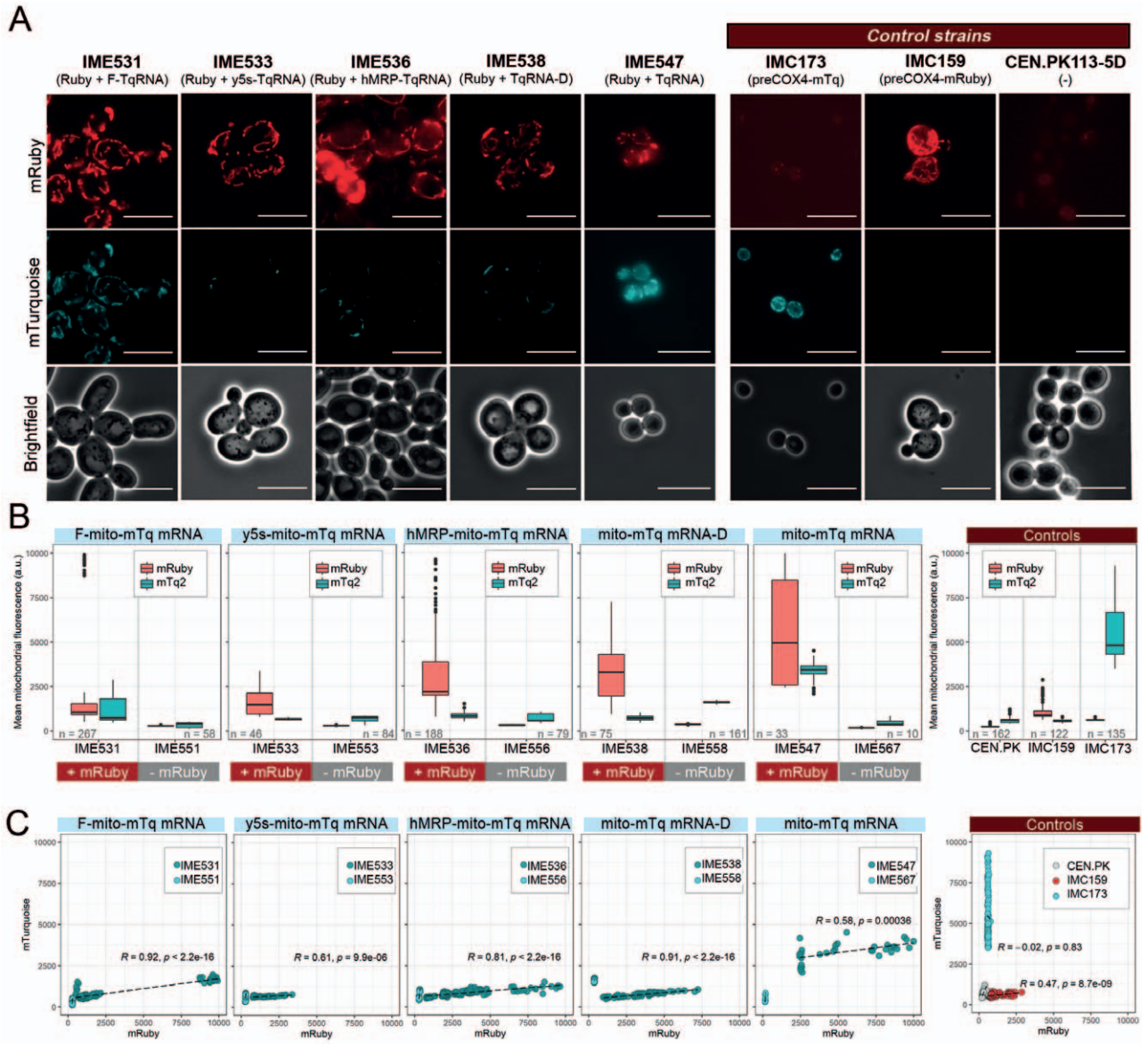


Figure 5.2. Analysis of colocalization of red (mRuby2) and blue (mTq2) fluorescence in strains expressing preCOX4-mRuby and/or mito-mTq2. Shown is a representative subset of data, the full dataset is shown in the supplementary data (Figure S5.5, Figure S5.6). A) Fluorescence microscopy images (1000 x) in red (mRuby), blue (mTurquoise) and brightfield images of strains expressing preCOX4-mRuby2 and mito-mTq2 fused to different mitochondrial RNA import signals (IME531-IME547), positive control IMC173, negative control IMC159 and parental strain CEN.PK113-5D. Scale bars represent 10 µm.
B) Distribution of the mean mitochondrial fluorescence in the different strains in the mRuby2 (red) and mTurquoise2 (blue) channels. The distribution is presented as boxplots, where the line represents the median value, and the edges of the boxes represent the 25th and 75th percentile. Outliers are shown as dots. Each plot contains mito-mTq2 plus an import signal in the presence (IME531-IME547) or absence (IME551-IME567) of preCOX4-mRuby2. C) Dot plots showing the correlation of the mean measured pixel intensity in the mRuby channel versus the measured mean pixel intensity in the mTurquoise2 channel. A linear trendline is shown as dashed line, R represents the Pearson correlation coefficient, p represents the p-value of the calculated correlation, a $p < 0.05$ is deemed significant.

acid sequence similarity with mTq2, but low similarity to mRuby2, and antibodies with affinity to GFP could therefore be used to specifically detect mTq2 in Western blots (Figure S5.8). Mitochondria were isolated from of a subset of strains that expressed mito-mTq2 mRNA only (IME551 – IME567), or mito-mTq2 mRNA in combination with the mRuby2 protein (IME531 – IME547), as well as the positive and negative control strains IMC159 (preCOX4-mRuby2 only), IMC173 (preSU9-mTq2 only) and CEN.PK113-7D (no fluorescent proteins expressed). The anti-GFP probe gave no signal in the negative control strains (lanes 1 and 2) and displayed the expected 26,9 kDa band for the positive control in which mTq2 is translated in the cytosol before mitochondrial import (lane 8). The double mito-mTq2 and preCOX4-mRuby2 strains displayed three bands, the 26,9 KDa corresponding to the mitochondrially translated mito-mTq2, and two smaller bands presumably corresponding to the truncated, cytosolic translated variants (lanes 3 to 7, Figure 5.3). The signal was particularly strong for strain IME533 in which the mito-Tq2 mRNA was fused in 5' to the y5S import signal. No signal was observed for full length mito-mTq2 in strains devoid of preCOX4-mRuby2 (lanes 9 to 13, Figure 5.3), but bands around 15 kDa, probably corresponding to cytosolic translated mito-Tq2, were visible.

Altogether, the localization of blue fluorescence to mitochondria and the detection of full size mito-mTq2 proteins in mitochondrial extracts suggested the presence of functional mTq2 in mitochondria, and thereby the mitochondrial import of its corresponding mRNA. The absence of blue fluorescence signal and of full size mTq2 in strains expressing only mito-mTq2 RNA and not preCOX4-mRuby2 suggested that mRuby2 expression might somehow facilitate mito-mTq2 RNA import to mitochondria.

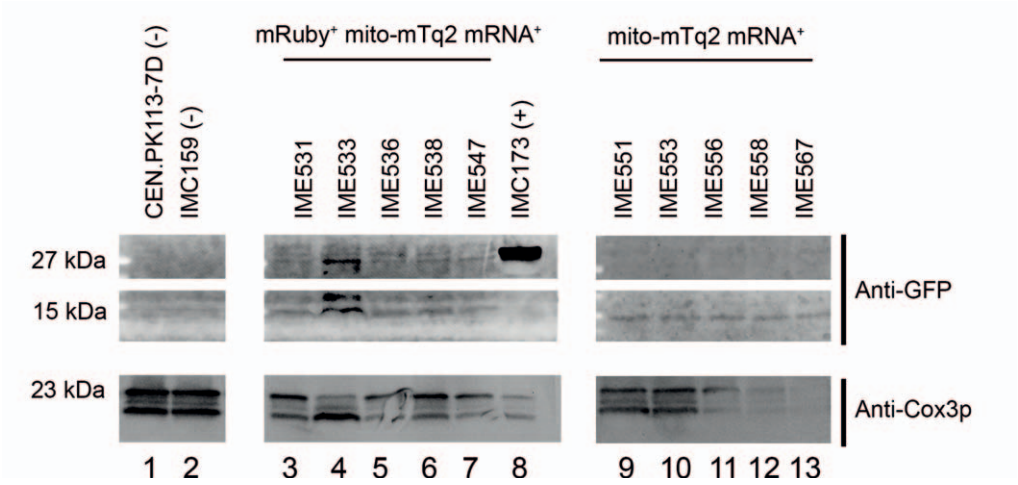


Figure 5.3. Western blot analysis of mTurquoise2 expression in strains expressing mito-mTq2 mRNA in the presence (IME531-547) and absence (IME551-567) of the of the mitochondrially localized mRuby2 protein. Negative controls (-) include CEN.PK113-7D (no fluorescent protein) and IMC159 (preCOX4-mRuby2), the positive control IMC173 (+) expressed preSU9-mTq2. Mito-mTq2 mRNA fused to different RNA import signals were tested - IME531/551: 5' F-stem, IME533/553: 5' y5S, IME536/556: 5' hMRP, IME538/558: 3' D-loop, IME547/567: no signal. The top two rows indicate bands observed on a blot treated with a fluorescent anti-GFP antibody. The lower row shows the same strains treated with an anti-Cox3p imaged using chemiluminescence. Cox3p is a mitochondrial protein used as loading control. Approximate protein sizes of the observed bands in kDa are indicated on the left. Each lane was loaded with 25 µg protein, except for IMC173 (10 µg protein). Expected protein sizes are 26.9 kDa for mTq2 and 30 kDa for Cox3p, but earlier experiments show that this protein runs lower in isolated mitochondria [45]. Uncropped blots are shown in SI figures (Figure S5.9)

Exploring the dependency of mitochondrial RNA translocation on overexpression of proteins localized to mitochondria

Mitochondrial import of $tRNA^{Lys}$(CUU) occurs through the TIM/TOM complexes, the same complexes through which proteins are imported into the mitochondria [18, 22, 70]. It is therefore not unreasonable to assume that high level expression and import to mitochondria of preCOX4-mRuby2 might affect the import of RNAs that are also imported via the TIM/TOM complexes. To determine whether this effect was specific to preCOX4-mRuby2, another protein, acetylornithine aminotransferase (Arg8p), was targeted to mitochondria

in combination with the mito-mTq2 mRNA. Acetylornithine aminotransferase is a mitochondrial protein, encoded by *ARG8* in the nucleus, that catalyses an essential step in arginine biosynthesis [71]. Deletion of *ARG8* in a *Δcar2* background causes stable arginine auxotrophy [42], which can be complemented by targeting the Arg8 protein to the mitochondrial matrix [72, 73]. To ensure the conditions were exactly the same, apart from the targeted protein, a variant of Arg8p was used in which the native Arg8-MTS was replaced by the preCOX4-MTS (IMC250, [42]). PreCOX4-Arg8p was, just like preCOX4-mRuby2, expressed on a centromeric (single-copy) plasmid in an *Δarg8Δcar2* strain together with a subset of constitutively highly expressed mito-mTq2 mRNA variants.

While growth in the absence of arginine confirmed the mitochondrial location of preCOX4-Arg8p, analysis of the constructed strains was complicated by the absence of a fluorescent marker for mitochondria localization. A fluorescence intensity threshold was set based on the fluorescence intensity of the non-fluorescent strains expressing mito-mTq2 with no auxiliary proteins (IME551 - IME567, Figure 5.2B). The mean fluorescence of sub-cellular particles expressing fluorescence higher that this threshold (1500) was determined in the negative control strain IMC250 *(Δarg8Δcar2*, complemented with preCOX4-*ARG8*), and in strains that co-expressed preCOX4-Arg8p and mito-mTq2 mRNA with different RNA import signals (IME797 – IME802). Strains expressing any form of the mito-mTq2 mRNA did not show a significant increase in fluorescence (ANOVA one-way analysis, $p > 0.05$, Figure 5.4A). As some outliers with an increased fluorescence intensity were observed, especially in strain IME800, the fluorescent signal was also plotted against the size of the fluorescent area, to select for cells that contained bright fluorescent spots with an area corresponding to the mitochondrial volume (0 – 15 μm², Figure 5.4B). Qualitative analysis of microscopy images of these cells did not show fluorescent structures reminiscent of mitochondria but rather vague spots that were also visible in

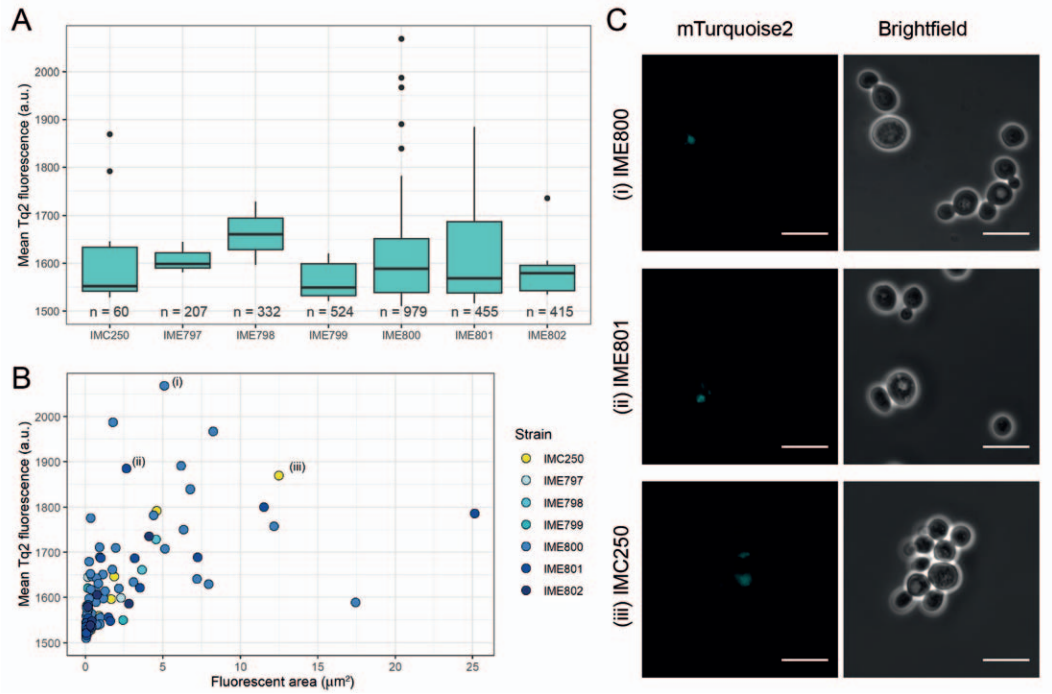


Figure 5.4. Impact of overexpression of the mitochondrial Arg8p on mito-mTq2 mRNA import and expression. IMC250 is the non-fluorescent control strain (*Δarg8 Δcar2*, plasmid with preCOX4-Arg8p), IME797-IME802 represent IMC250 containing a plasmid expressing with the mito-mTq2 mRNA fused to different RNA import signals (797: 5' F-stem, 798: 5' y5S, 799: 5' hMRP, 800: 3' D-loop, 801: 5' tRK2-tr93, 802: no signal). A) Distribution of the mean fluorescence in the mTurquoise2 (blue) channel, determined from microscopic analysis and measuring mean pixel intensities in the blue channel with a pixel intensity of > 1500. The distribution is presented as boxplots, where the line represents the median value, and the edges of the boxes represent the 25th and 75th percentile. Outliers are shown as dots. n indicates the number of analyzed cells. B) Mean fluorescence in the mTurquoise2 (blue) channel of each cell plotted against the size of the fluorescent area. C) Fluorescence microscopy images (1000x) in blue (mTurquoise2) and brightfield images of cells indicated with i, ii, and iii in panel B. Scale bars represent 10 μm.

the negative control (Figure 5.4C). Therefore, targeting of preCOX4-*ARG8* to the mitochondria did not visibly enhance RNA import, suggesting that this effect might be dependent on the targeted protein.

Bioreactor evolution and cultivation of strains expressing mito-*ARG8* mRNA

Current understanding of mitochondrial RNA import in yeast is limited to the import mechanism of $tRNA^{Lys}$(CUU) [22]. Although several $tRNA^{Lys}$(CUU)-derived signals were used in this study (Table 5.2), none did measurably and consistently improve mRNA import. The import mechanism of the other tested signals is not fully elucidated, and fluorescence microscopy and Western blotting indicated that even in absence of an RNA import signal, RNA can still be imported and translated. RNA import of mRNA-sized RNA might therefore be a stochastic rather than a selective process, although it may be improved by additional protein targeting to the mitochondria. As sufficient knowledge on RNA import mechanisms to rationally engineer mRNA import in mitochondria is lacking, an adaptive laboratory evolution approach (ALE) was pursued, which is particularly efficient to identify the molecular basis of complex phenotypes [74]. Evolving cells for improved mitochondrial RNA import requires a condition where RNA import is essential for survival, or at least gives a strong selective advantage. This condition can be met in an *Δarg8Δcar2* arginine auxotroph [42], if the only source of the essential Arg8p is supplied in the form of an mRNA that must be imported and translated in the mitochondria to rescue arginine auxotrophy. Eleven different 5' RNA import signals were cloned in front of the mito-*ARG8* mRNA construct, omitting the signal in one construct, in a high-copy vector under control of a strong constitutive promoter. The twelve constructs were transformed in a *Δarg8Δcar2* arginine auxotroph (IMX2796). IMX2796 constitutively expressed preCOX4-mRuby2 through a genomic integration. None of the strains were readily able to grow in the absence of arginine (Figure S5.10), indicating that RNA import or translation of the mito-*ARG8* mRNA did not occur or was not efficient enough to sustain arginine requirements for growth.

To improve RNA import in these strains, an ALE experiment was designed. A chemostat setup was used, in which nitrogen was the limiting element, continuously supplied in the form of ammonium sulfate and arginine. Unable to synthesize arginine from ammonium, the constructed strain solely relied on the supplied arginine for growth. It was expected that through ALE, strains with mutations facilitating mito-*ARG8* mRNA import would occur, thereby enabling arginine synthesis from the supplied ammonium. This would result in an increased biomass yield of these mutants, which would gradually take over the culture. Ethanol was used as carbon source, to increase mitochondrial mass, retain a mitochondrial membrane potential and prevent loss of mitochondrial DNA [75, 76].
Two bioreactors were inoculated with a pre-culture containing a mixture of the twelve strains present in equal cell number. The cultures were maintained at a dilution rate (equivalent to the specific growth rate) of 0.1 h^{-1}. The theoretical biomass concentration defined by the limiting arginine supply in the medium (60 mg L^{-1}) was around 2 g L^{-1} [77]. Full consumption of arginine and ammonium would result in a maximum biomass concentration of 12 g L^{-1}, with complete consumption of the supplied ethanol (435 mM). The initial batch phase reached a biomass concentration of 3 g L^{-1}, which was higher than expected and likely due to arginine carry-over from the preculture. Then, the cultures were characterized during the first 600 hours (ca. 86 generations) by a relative steady-state with an average of 1.5 g L^{-1} biomass, and oscillating carbon dioxide and oxygen concentrations. Such oscillations, distinct from fast metabolic oscillations [78], have been previously reported for chemostat cultures of *S. cerevisiae* and might reflect cellular stress due to the limited arginine supply and periodic shifts in intracellular arginine pools [79-83]. After approximately 750 hours (*ca.* 107 generations), the oscillations disappeared while biomass concentration increased to ca. 2.4 g L^{-1}. Concomitantly, carbon dioxide production increased by 40 – 50 % and residual ethanol concentration decreased by 10 % (Figure 5.5A,B).

The ability of the strains to grow in the absence of arginine outside the bioreactors was tested by plating. No growth was observed on arginine-free medium for cell samples originating from bioreactor A, however cell samples from bioreactor B yielded approximately 100 colony-forming units (CFU) per 10^7 cells plated after three weeks of growth on arginine-free medium. These results suggested that a fraction of the cell population in bioreactor B had acquired the ability to synthetize arginine. To increase arginine synthesis, and thereby

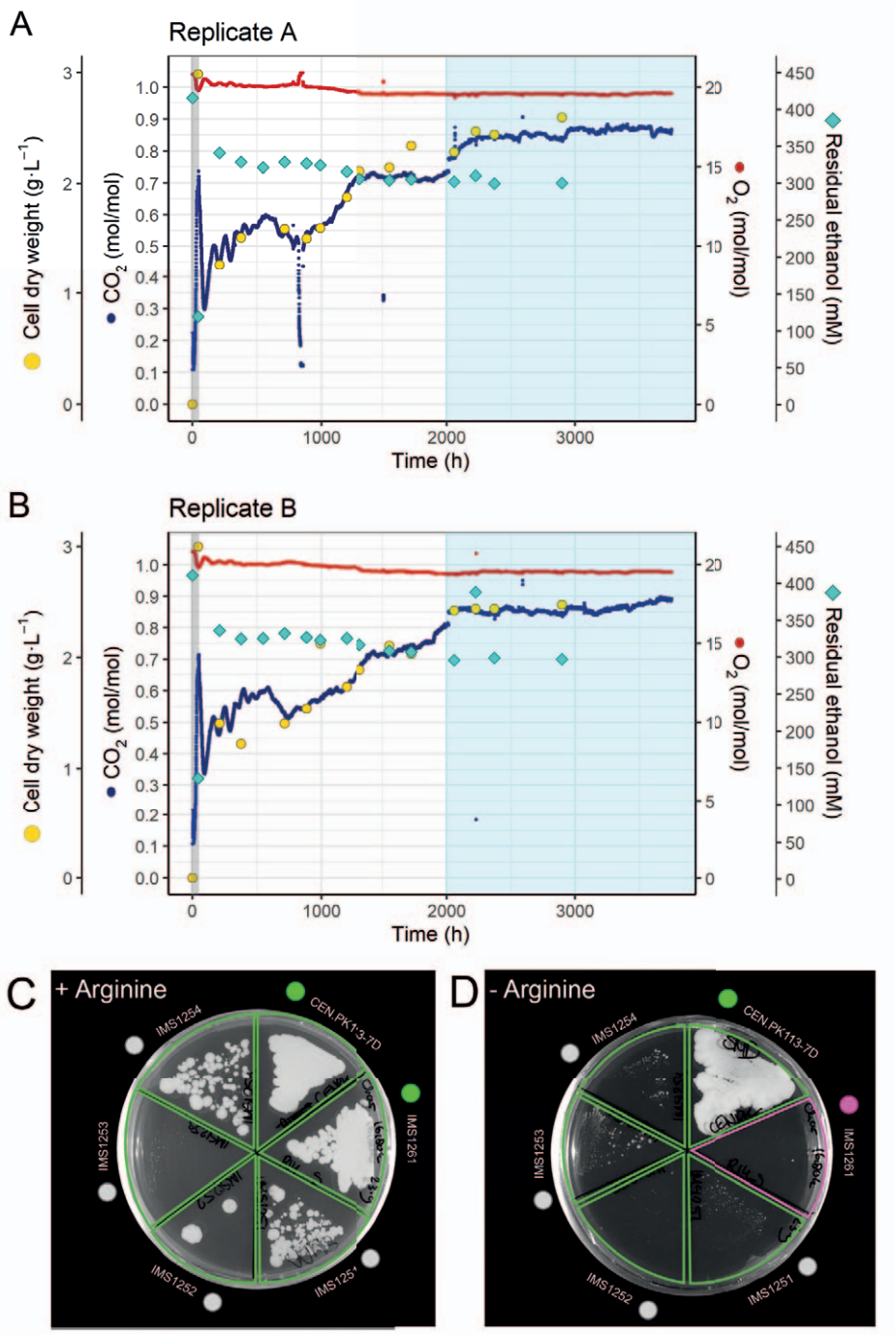


Figure 5.5. Prolonged chemostat culture of *S. cerevisiae* to evolve towards mito-*ARG8* mRNA import in mitochondria. A,B) Profiles of two replicate bioreactors inoculated with a mixture of IMC224-IMC235, arginine auxotrophic strains expressing mito-*ARG8* mRNA and the preCOX4-mRuby2 protein. The chemostats were operated at a dilution rate of 0.1 h^{-1}, under nitrogen (arginine) limitation and ethanol as source carbon and energy source. The grey area represents the batch phase, the blue area highlights the phase with relatively steady biomass concentration. C,D) Growth of four single-colony isolates IMS1251-IMS1254 from bioreactor replicate B on medium with (C) and without (D) arginine, along with control strain CEN.PK113-7D (prototroph) and the population of bioreactor B at the end of the batch phase (IMS1261).

growth rate, 1/10th of the continuous culture volume in bioreactor B was used to start sequencing batch reactors [74] with arginine-free medium. With the strong selection pressure caused by arginine absence, the yeast culture was expected to evolve towards more efficient mito-*ARG8* import or arginine synthesis and thereby a faster specific growth rate during successive batching. Six successive empty-refill batches were performed over a 27-days period, and a ca. 20 % increase in biomass concentration was observed in each batch culture. However, as the specific growth rate (0.003 ± 0.001 h^{-1}), CO_2 production and O_2 consumption did not visibly improve (Figure S5.11) in this time frame, evolution of these extremely slowly growing cultures was discontinued.

Characterization of mutants able to grow in the absence of arginine.

To identify the molecular mechanisms involved in the acquisition of arginine prototrophy during the ALE experiment, four single colony isolates able to grow in the absence of arginine were isolated from bioreactor B (strains IMS1251 to IMS1254). As observed for the whole evolved population, the four strains grew poorly on plates with arginine-free medium with glucose and the growth of IMS1253 was also impaired in the presence of arginine (Figure 5.5C,D). IMS1254 was selected for physiological characterization and grown in arginine-free batch cultures with glucose or ethanol as sole carbon and energy source. Although the strain was able to double in cell count approximately once in 500 hours in both media (an approximately 300-fold reduction in doubling time compared to wild-type *S. cerevisiae*), the optical density remained extremely low as cell volume decreased, and changes in metabolic activity (consumption of oxygen and glucose, production of carbon dioxide, etc.) as well as biomass concentration were not measurable (Figure S5.12, Figure S5.13).

The poor growth observed on arginine-free medium may be the result from the occurrence of mutations enabling mitochondrial import of mito-*ARG8* mRNA and arginine biosynthesis, but it may also be caused by an increased cellular efficiency for growth in arginine-poor media. Which of the two mechanisms is involved was tested by sequencing the genome of the four evolved strains, as well as the two evolved populations from bioreactors A and B (named IMS1248 and IMS1249, respectively). A set of 38 non-synonymous mutations were found in one or more of the evolved isolates (SI file 1). Considering that the population of bioreactor A did not evolve the ability to grow in arginine-free medium, mutations that were shared between IMS1248 and the evolved isolates were considered as not relevant for acquisition of arginine prototrophy or could have occurred during preculture growth and were therefore discarded. The remaining five mutated genes were *ASG1* and *ECM21*, each mutated in a single evolved strain, *TIF3* and *YOR389w*, each harboring the exact same mutation in two evolved strains, and *UBR2* displaying the same mutation in all four evolved strains (Table 5.3). Four out of these five proteins are involved in protein turn-over, either via synthesis (*TIF3* encodes the translation initiation factor eIF-4B and *ASG1* is a transcriptional regulator) or degradation through the proteasome (*ECM21* and *UBR2*). In all four strains, the *UBR2* STOP codon was replaced by a tyrosine, thereby extending the C-terminus by 27 extra amino acids (Table 5.3). Ubr2 is a cytoplasmic E3 ubiquitin-protein ligase involved in proteasome regulation via the degradation of Rpn4p, transcriptional activator of proteasome genes. Loss of activity of Ubr2p might therefore stimulate the transcription of proteasome subunits [84, 85] but it might also have adverse effects by decreasing the Ubr2p-mediated degradation of unfolded and misfolded proteins.

Ecm21p is a ubiquitin ligase adaptor for Rsp5p, a NEDD4 family E3 ubiquitin ligase [86, 87] involved in protein degradation and regulates starvation- and substrate-induced ubiquitin-dependent endocytosis of specific plasma membrane amino acid transporters [88]. In strain IMS1253, a 2-bp insertion occurred in this gene (SI file 1), resulting in early termination after 198 amino acids of the otherwise 1119 amino acid-long protein. Interestingly, IMS1248 carried an amino acid substitution ($Leu_{518}Ser$) in the strongly conserved HECT-domain of Rsp5 required for ubiquitylation activity [89], suggesting that alteration of protein degradation in general, or more specifically of amino acid permeases (including arginine), might give a selective advantage in arginine-limited cultures. Although the mutations in *TIF3*, *ASG1* and *YOR389w* might be relevant for arginine prototrophy, their physiological role in this context remains to be elucidated.
The batches were started with a mixture of strains, each with a different putative RNA import signal attached to the mito-*ARG8* mRNA. All four evolved isolates solely contained a plasmid containing the 72-base pair tr93-tRK2 import signal, a modified version of the full $tRNA^{Lys}$(CUU) tRK2 that was described to be imported

Table 5.3. Non-synonymous mutations found in populations IMS1248 and IMS1249 and single colonies IMS1251 - IMS1254. Mutations that were found in both IMS1248 and IMS1249 were excluded as these could originate from the preculture. A full list of all mutations can be found in SI file 1.

Location	Position	Systemic name	Gene name	Description	Subcellular localization	Mutation	Effect on protein	Found in strains:
CHR II	28040	*YBL101C*	*ECM21*	Alpha-arrestin, ubiquitin ligase adaptor for Rsp5p; regulates starvation- and substrate-induced Ub-dependent endocytosis of select plasma membrane localized amino acid transporters	Cytosol	Insertion (2 bp)	Early termination (198 aa)	IMS1253
CHR V	412435	*YER125W*	*RSP5*	NEDD4 family E3 ubiquitin ligase	Cytosol	Leu518Ser	Amino acid substitution	IMS1248
CHR IX	99639	*YIL130W*	*ASG1*	Activator of stress genes 1, Zinc cluster protein proposed to be a transcriptional regulator; regulator involved in the stress response; regulates utilization of fatty acids and accumulation of lipids	Nucleus	Insertion (9 bp)	Insertion 3x Asn	IMS1249, IMS1252
CHR XII	174490	*YLR024C*	*UBR2*	E3 ubiquitin-protein ligase component of the Mub1p-Ubr2p-Rad6p ubiquitin ligase complex required for the ubiquitination and degradation of Rpn4p	Cytoplasm	STOP1283Tyr	Elongation (27 aa)	IMS1249, IMS1251, IMS1252, IMS1253, IMS1254
CHR XV	1069766	*YOR389W*		Putative protein of unknown function; expression regulated by copper levels		Asn437Asp	Amino acid substitution	IMS1253, IMS1254
CHR XVI	871051	*YPR163C*	*TIF3*	Translation initiation factor eIF-4B; RNA recognition motif containing single-stranded RNA binding protein that possesses RNA annealing and strand-exchange activities	Cytoplasm	Deletion (1 bp)	Early termination (348 aa)	IMS1249, IMS1252, IMS1253

in mitochondria [90]. However, since many mutations are shared between the four evolved isolates, it may also indicate the strains share a common ancestor, rather than indicating tr93-tRK2 is an optimal RNA import signal. No mutations that could mitochondrial enhance mRNA import or translation were detected in the plasmid sequences.

Altogether these results suggested that the evolved strains became more efficient for growth under low arginine supply, potentially by altering protein turn-over. The sequencing data did not provide information on alternative routes for arginine synthesis, nor did they lead to the identification of potential candidates for improvement of RNA import to mitochondria.

Proteome analysis of the evolved mutants does not indicate mitochondrial import of mito-Arg8 mRNA.

LC-MS shotgun proteomics of isolated mitochondria of the evolved strains was performed to verify if Arg8p was translated and localized to mitochondria. The evolved strains were grown in the presence of arginine with ethanol as carbon source, as biomass concentrations reached without arginine were too low for mitochondria isolation. IMS1253 was excluded from this analysis as its growth was poor even in the presence of arginine. Unfortunately, none of the peptides of the Arg8p protein were detected in the proteome of the isolated mitochondria of evolved strains, indicating that the mito-*ARG8* mRNA was most probably not imported or translated in mitochondria.

Nevertheless, measured changes in proteome could give insight into the physiological changes caused by ALE that allowed the strains to grow in the absence of arginine without active Arg8p. The proteomes of strains IMS1251, IMS2152 and IMS1254, were compared to the non-evolved starting culture IMS1261. A total of 927 proteins was detected across all samples, with various degrees of enrichment for mitochondrial proteins in each strain. IMS1251, IMS1252 and IMS1254 contained 49 %, 68 % and 53 % mitochondrial protein, respectively. The non-evolved parental population IMS1261 contained 71 % mitochondrial protein (Figure S5.14). Changes in the proteomes of the evolved mutants were determined by normalizing protein abundance over the protein abundances found in the unevolved starting culture IMS1261. IMS1251 and IMS1254 showed a similar response, with *ca.* 330 proteins that were significantly more abundant ($p < 0.05$) and approximately 200 proteins that were less abundant compared to the non-evolved parental population (Figure S5.15, SI file 2). The changes observed in IMS1252 were less extensive, with 251 and 121 proteins that had higher and lower abundance as compared to the non-evolved parental population, respectively. As common feature, a total of 205 proteins were more abundant, and 66 proteins were less abundant in all three strains as compared to the non-evolved control strain (Figure S5.15). Most of these proteins were not mitochondrial.
Analysis of the mitochondrial proteome did not reveal remarkable relative changes in general proteome allocation during evolution (Figure S5.16). Among the ca. 1000 mitochondrial proteins, as little as six and twenty-five were more and less abundant, respectively, in all strains compared to the unevolved IMS1261 (Figure 5.6A,B). A GO-term analysis on the proteins more abundant in all three strains revealed that half of these proteins were involved in aminoacylation of tRNA (*ALA1*, *VAS1* and *ISM1*, Figure 5.6C, SI file 2). The less abundant proteins were largely involved in the mitochondrial electron transport chain. Four genes were involved in cytochrome *c* oxidase activity (all *COX*-genes), six proteins play a role in oxidative stress (*PIM1*, *COX20*, *CCP1*, *SOD2*, *POS5*, *TRX3*) and eight proteins were involved in respiration (*COX*-genes, *KGD2*, *ETR1*, *MDH1*, *LSC2*, Figure 5.6D, SI file 2). Remarkably, 78 of the less abundant proteins were shared between IMS1251 and IMS1254 (Figure 5.6B), of which 30 were components of mitochondrial ribosomes, while an average of 50 proteins was involved in metabolism of peptides and organic nitrogen (Figure S5.17). This could hint at an altered (nitrogen-containing) amino acid metabolism or shift in allocation of amino acids.

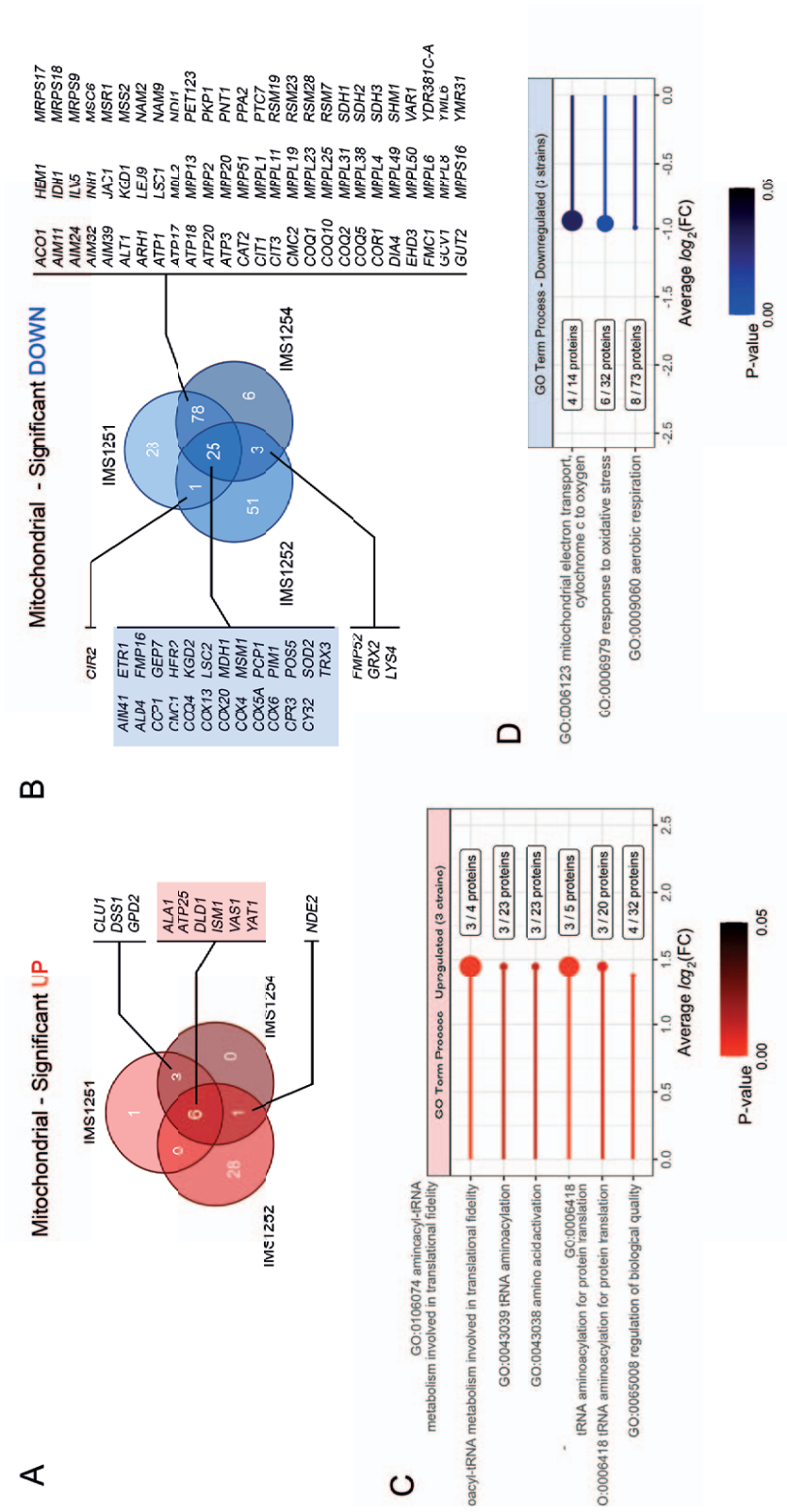


Figure 5.6. Proteome response of mitochondrial proteins detected in IMS1251, IMS1252 and IMS1254 compared to unevolved parental population IMS1261. A) Venn diagram of the significantly more abundant mitochondrial proteins of each strain. Numbers indicate the number of proteins more abundant in one, two or all three strains. Genes more abundant in more than one strain are listed, genes more abundant in all three strains are indicated in red. B) Venn diagram of the significantly less abundant mitochondrial proteins of each strain. Numbers indicate the number of proteins more abundant in one, two or all three strains. Genes less abundant in more than one strain are listed, genes less abundant in all three strains are indicated in blue. C) GO-term analysis of six proteins that were more abundant in all strains (indicated in red in (A)) and D) GO-term analysis of 25 proteins that were less abundant in all strains (indicated in blue in (B)). Genes were classified according to Gene Ontology (GO) terms of the types "Process". The length of each bar indicates the average measured log_2(fold change, (FC)) of all genes associated with the GO-term. The size of the dot indicates the enrichment strength (number of detected proteins in each term/number of proteins of each term in the reference list), which is also indicated next to each bar. The color of each bar indicates the significance of the enrichment, all were below p = 0.05 (Bonferroni corrected).

5.4 Discussion

Mitochondrial import of RNA remains an elusive topic to date, and characterization is limited to the import of short tRNA or gRNA (100 - 200 nucleotides). gRNA localization is often demonstrated through depletion of mtDNA by targeting Cas9 and a gRNA to the mitochondrial matrix. Although several mitochondrial CRISPR-studies demonstrated a 25 – 50 % reduction of mtDNA abundance in the presence of a Cas9 protein and a gRNA [28, 91, 92], there is no direct evidence of gRNA targeting to the mitochondrial matrix. A study by Schmiderer *et al.* [34] uncovered that a mitochondrially targeted Cas9 base-editor had a 1 % editing efficiency on the same locus of the mtDNA, regardless of the sequence of the targeted gRNA. This data suggested that Cas9 can have off-target effects on the mtDNA regardless of the presence of the gRNA. mtDNA is intrinsically highly variable in copy number. A DSB in the mtDNA, through e.g., off-target efficiency of Cas9, can lead to mutations in mtDNA or trigger maintenance systems such as mitophagy [33, 93, 94], resulting in a reduced number of mtDNA copies. Depletion of mtDNA is therefore not a conclusive readout of gRNA targeting. Other studies investigating RNA import often use a combination of mitochondrial subfractionation and subsequent qPCR or Northern blot to demonstrate mitochondrial RNA localization. However, cytosolic RNA adheres to the mitochondrial surface [10, 34, 95-98], and isolated mitochondria, even in highly purified or mitoplast form, still contain circa 20 - 80 % contamination by cytosolic material [95, 99]. Therefore, detection of mRNA import is challenging with these sensitive methods as they do not explicitly demonstrate RNA import, but only RNase protection of the targeted RNA by association to the mitochondrial membrane or mitochondrial proteins [33].

With these challenges in mind, it is required to establish a different method to reproducibly demonstrate RNA import, without relying on Cas9 activity and prevent false-positives caused by cytosolic contamination. In this study, this was attempted by targeting a mitochondrially recoded mRNA to the mitochondrial matrix. Mitochondrially recoded genes can only be translated in the mitochondria, offering a powerful screening method of mRNA import to mitochondria [73]. Using this method, mRNA import in the mitochondria was detected, however, the presence or absence of an RNA import signal did not have a measurable effect on mRNA import in the mitochondria. Most of these signals have been developed and tested *in vitro* for the import of short RNA molecules on isolated mitochondria and may therefore not be suitable for (*in vivo*) long mRNA import [23, 24, 26-28]. Random import of RNA regardless of its sequence is highly unlikely, as it would result unspecific RNA accumulation in the mitochondria. Another unexpected result was the detection of mito-mTq2 fluorescence and full-length protein only when mRuby2 was co-expressed and translocated to the mitochondria. As the TIM and TOM protein import complexes are likely involved in RNA import to mitochondria [22], high nuclear expression and mitochondrial import of an additional protein might affect RNA import. However, approximately 1,000 proteins are natively imported into mitochondria, and are sorted either in the mitochondrial membrane or in the matrix based on their localization signal [2, 100], causing intensive protein trafficking across the mitochondrial membrane in active cells. It is therefore unlikely that expressing and importing yet another protein affects RNA import. This hypothesis was additionally experimentally disproved by the lack of mito-mTq2 mRNA import in strains expressing Arg8p using the same expression system and MTS as mRuby2.

Alternatively, the positive effect of mRuby2 on mito-mTq2 mRNA import might be due to a specific feature of mRuby2. Mitochondrial co-import of RNA and specific proteins has been described in multiple occasions in yeast and mammals, and although the mechanisms are often disputed [10, 24, 25, 34, 101-103], it is possible that mRuby2 possesses intrinsic characteristics that aid mRNA import. For example, relatively small proteins like preCOX4-mRuby2 (29 kDa) are more easily imported than large proteins, such as the 47 kDa protein Arg8 [104]. Also, the protein carries a net positive charge ($pI_{preCOX4\text{-}mRuby2}$ = 8.77, calculated using ExPaSY ProtParam [105]) which could have a depolarizing effect on $\Delta\Psi$ and may therefore aid in transport of negatively charged RNA molecules across the mitochondrial membrane [104, 106], although its charge is comparable to preCOX4-Arg8 ($pI_{preCOX4\text{-}Arg8}$ = 8.03). The preCOX4 mitochondria localization signal used for mRuby2 and Arg8 is derived from the cytochrome *c* oxidase subunit 4 (Cox4). Cox4 is localized in the inner mitochondrial membrane and its import is dependent on the mitochondrial membrane potential $\Delta\Psi$ [107-109]. The exact submitochondrial localization of mRuby2 and Arg8 fused to the *COX4*-MTS is not

known, but their chemical and physical properties might lead to a different localization and effect on RNA import into mitochondria. Nevertheless, mRuby2 does not have any known RNA-binding functions, which makes its role in mTq2-mRNA import rather unlikely. Another important observation was the stochasticity and poor intensity of mito-mTq2 fluorescence between mitochondria and between cells, indicating that import of mito-mTq2 was a rare event. Last but not least, it cannot be excluded that the assumption of successful RNA import to mitochondria was based on technical artefacts, despite positive results shown both by fluorescence microscopy and by Western Blotting.

Potential artefacts introduced by fluorescence microscopy were avoided by expressing mito-*ARG8* as mitochondrially targeted mRNA. High expression of the mRNA from a multicopy plasmid with a strong constitutive *TEF1*-promoter was not sufficient to complement the metabolic requirement of *Δarg8* mutants auxotrophic for arginine (Figure S5.10). This result suggested either that *ARG8* mRNA was not readily imported to mitochondria or that Arg8p expression, if present, was too low to sustain growth requirements. The latter hypothesis served as basis for the adaptive laboratory evolution approach. The ALE experiment was carefully designed with a strong selection pressure favoring mutants with enhanced mitochondrial import of mito-*ARG8* mRNA. However, the evolved strains did not express mito-Arg8 in their mitochondria, indicating mito-*ARG8* mRNA was not imported. In line with this result, whole genome sequencing and mitochondrial proteome analysis of the evolved strains did not reveal potential mechanisms enhancing RNA import to mitochondria. Regardless, the evolved strains were able to bypass the arginine biosynthetic pathway, but only at extremely low rates (Figure S5.12, Figure S5.13). The mechanism behind this is not exactly clear, but WGS data pointed at a change in protease activity, as a mutation in *UBR2*, a cytoplasmic E3 ubiquitin-protein ligase required for feedback regulation of the 26S proteasome [84, 110, 111], was detected in quadruplicate in the evolved strains. The mutation caused extension of the *UBR2* ORF, potentially affecting the regulatory mechanism of Ubr2p. This may cause an altered 26S proteasome metabolism [85, 112], which could ultimately lead to increased protein turnover, proteasome turnover [112], or conversely, the stabilization of proteins [113], all reducing the demand for arginine. Recycling of the available intracellular arginine cannot constitute growth for multiple generations, coinciding with the observation that the evolved mutants were not able to double more than once (Figure S5.12). Interestingly, the 26S proteasome also interacts with the import of mitochondrial $tRNA^{Lys}$(CUU) and has a positive effect on the import of the tRNA by degradation of several cytosolic tRNA-retention factors [114]. Although it is possible that altered proteasome homeostasis was a step towards improved mitochondrial RNA import, additional mutations were still required to attain import.

Proteomics indicated a reduction in expression of mitochondrial respiratory proteins in the mutants isolated from the chemostat evolution (Figure 5.6). Such an effect was previously described for prolonged cultivation of *S. cerevisiae* in aerobic glucose-limited chemostats at a dilution rate below the μ_{max}, where the slow uptake of glucose and glycolytic flux resulted in a strong reduction of glycolytic enzymes, thereby preventing wasteful allocation of resources to the synthesis of superfluous proteins [115]. A similar mechanism likely occurred in the present ALE experiments, which occurred at a reduced growth rate, approximately 75 % of the μ_{max}, in which respiration functions below its maximum capacity. Proteome analysis indeed revealed reduced levels of proteins involved in respiration in the evolved strains, while the specific oxygen uptake rate remained unchanged.Even if not successful in the present study, the designed ALE strategy has the potential to improve RNA import to mitochondria. An increase in biomass yield was observed during prolonged chemostat cultivation, which most likely resulted from cellular optimization to arginine-poor conditions, and not from improved mito-*ARG8* import. Establishing a baseline response of arginine auxotrophic control strains devoid of mitochondrial *ARG8* mRNA import would enable to disentangle 'natural' yeast responses to these culture conditions from effective import of mito-*ARG8* mRNA.

Despite the implementation of innovative approaches based on mitochondria-specific translation and ALE, import of mRNA-sized RNAs was not achieved in the present study. The pitfalls of standard methods to demonstrate RNA import of non-native mRNA into mitochondria, encountered in this study as well as other studies in RNA import, strongly advocate to further develop innovative and more specific approaches to detect RNA import. The import of a long mRNA (up to 1272 nucleotides for *ARG8*) may be too ambitious, but a similar approach might be successful with shorter RNAs. Designing an ALE experiment with selection

pressure for the import of a short RNA in mitochondria is challenging, but not impossible. For instance, selecting for complementation of mitochondrially-encoded tRNA, upon removal of the native tRNA encoded by the mtDNA, might offer a powerful ALE strategy.

Acknowledgements

The authors thank Mark Bisschops for input in experiment design and assistance with FISH and Western Blot experiments, Anne Schwabe, Jeremie Capoulade and Michal Shemesh for performing confocal microscopy experiments, Marcel van den Broek for assistance in WGS data analysis and Dita Heikens for assistance in proteome preparation.

5.5 Supplementary data

Supplementary files SI_1_WGS and SI_2_Proteomics_GOterm can be accessed through the 4TU Research Data Database (https://data.4tu.nl/) under doi:10.4121/12375721-959a-4452-a8cb-dc44277bfe86 or via the qr code below:

Links to: https://data.4tu.nl/datasets/12375721-959a-4452-a8cb-dc44277bfe86

Sequences

DF-stem (retrieved from [23])

```
1 gcgcaatcgg tagcgcttcg agccccctac agggctcca
```

D-arm (retrieved from [23])

```
1 gcgcaatcgg tagcgc
```

F-arm (retrieved from [23])

```
1 gagcccccta cagggctctt
```

F1 (retrieved from [23])

```
1 ggttcgagcc ccctacaggg ctcca
```

tRK2-tr93 (retrieved from [23])

```
1  gccttgttag ctcagttggt agagcgttcg gcttttaacc gaaatgtcag gggttcgagc
61 cccctatgag gct
```

Human 5S rRNA (retrieved from [26])

```
1 ggcctggtta gtacttggat gggggaccgc caaggaatac cgggtg
```

Yeast 5S (Figure S5.3)

```
1 gaccgagtag tgtagtgggt gaccatacgc gaaactcagg tg
```

Human ribonuclease P stem (retrieved from [24])

```
1 tctccctgag cttcagggag
```

Yeast ribonuclease P stem (Figure S5.3)

```
1 actggggaac cagt
```

Human mitochondrial RNA processing ribonuclease stem (retrieved from [24])

```
1 agaagcgtat cccgctgagc
```

Yeast mitochondrial RNA processing ribonuclease stem (Figure S5.3)

```
1 gtacctctat tgcagggtac
```

tracrRNA (from pMEL13 [116])

```
1   gttttagagc tagaaatagc aagttaaaat aaggctagtc cgttatcaac ttgaaaaagt
61  ggcaccgagt cggtggtgc
```

Mito-mTq2

The sequence for mTq2 was retrieved from [117], and used for codon-optimization for yeast mitochondrial translation.

```
1    atggtatcaa aaggtgaaga attatttaca ggtgtagtac ctattttagt agaattagat
61   ggtgatgtaa atggtcataa attttcagta tcaggtgaag gtgaaggtga tgctacatat
121  ggtaaattaa cattaaaatt tatttgtaca acaggtaaat tacctgtacc ttgacctaca
181  ttagtaacaa cattatcatg aggtgtacaa tgttttgcta gatatcctga tcatatgaaa
241  caacatgatt tttttaaatc agctatgcct gaaggttatg tacaagaaag aacaattttt
301  tttaaagatg atggtaatta taaaacaaga gctgaagtaa aatttgaagg tgatacatta
361  gtaaatagaa ttgaattaaa aggtattgat tttaaagaag atggtaatat tttaggtcat
421  aaattagaat ataattattt ttcagataat gtatatatta cagctgataa acaaaaaaat
481  ggtattaaag ctaattttaa aattagacat aatattgaag atggtggtgt acaattagct
541  gatcattatc aacaaaatac acctattggt gatggtcctg tattattacc tgataatcat
601  tatttatcaa cacaatcaaa attatcaaaa gatcctaatg aaaaaagaga tcatatggta
661  ttattagaat ttgtaacagc tgctggtatt acattaggta tggatgaatt atataaataa
```

Mito-*ARG8*

Includuding *COX2*-UTL and dodecamer, underlined

```
1    agtattaaca tattataaat agacaaaaga gtctaaaggt taagatttat taaaatgttt
61   aaaagatatt tatcatcaac atcatcaaga agatttacat caattttaga agaaaaagct
121  tttcaagtaa caacatattc aagacctgaa gatttatgta ttacaagagg taaaaatgct
181  aaattatatg atgatgtaaa tggtaaagaa tatattgatt ttacagctgg tattgctgta
241  acagctttag gtcatgctaa tcctaaagta gctgaaattt tacatcatca agctaataaa
301  ttagtacatt catcaaattt atattttaca aaagaatgtt tagatttatc agaaaaaatt
361  gtagaaaaaa caaaacaatt tggtggtcaa catgatgctt caagagtatt tttatgtaat
421  tcaggtacag aagctaatga agctgcttta aaatttgcta aaaaacatgg tattatgaaa
481  aatccttcaa aacaaggtat tgtagctttt gaaaattcat ttcatggtag aacaatgggt
541  gctttatcag taacatgaaa ttcaaaatat agaacacctt ttggtgattt agtacctcat
601  gtatcatttt taaatttaaa tgatgaaatg acaaaattac aatcatatat tgaaacaaaa
661  aaagatgaaa ttgctggttt aattgtagaa cctattcaag gtgaaggtgg tgtatttcct
721  gtagaagtag aaaaattaac aggtttaaaa aaaatttgtc aagataatga tgtaattgta
781  attcatgatg aaattcaatg tggtttaggt agatcaggta aattatgagc tcatgcttat
841  ttaccttcag aagctcatcc tgatattttt acatcagcta aagctttagg taatggtttt
901  cctattgctg ctacaattgt aaatgaaaaa gtaaataatg ctttaagagt aggtgatcat
961  ggtacaacat atggtggtaa tcctttagct tgttcagtat caaattatgt attagataca
1021 attgctgatg aagcttttTT aaaacaagta tcaaaaaaat cagatatttt acaaaaaaga
1081 ttaagagaaa ttcaagctaa atatcctaat caaattaaaa caattagagg taaaggttta
1141 atgttaggtg ctgaatttgt agaacctcct acagaagtaa ttaaaaaagc tagagaatta
1201 ggtttattaa ttattacagc tggtaaatca acagtaagat ttgtacctgc tttaacaatt
1261 gaagatgaat taattgaaga aggtatggat gcttttgaaa aagctattga agctgtatat
1321 gcttaattat aataatattc ttaa
```

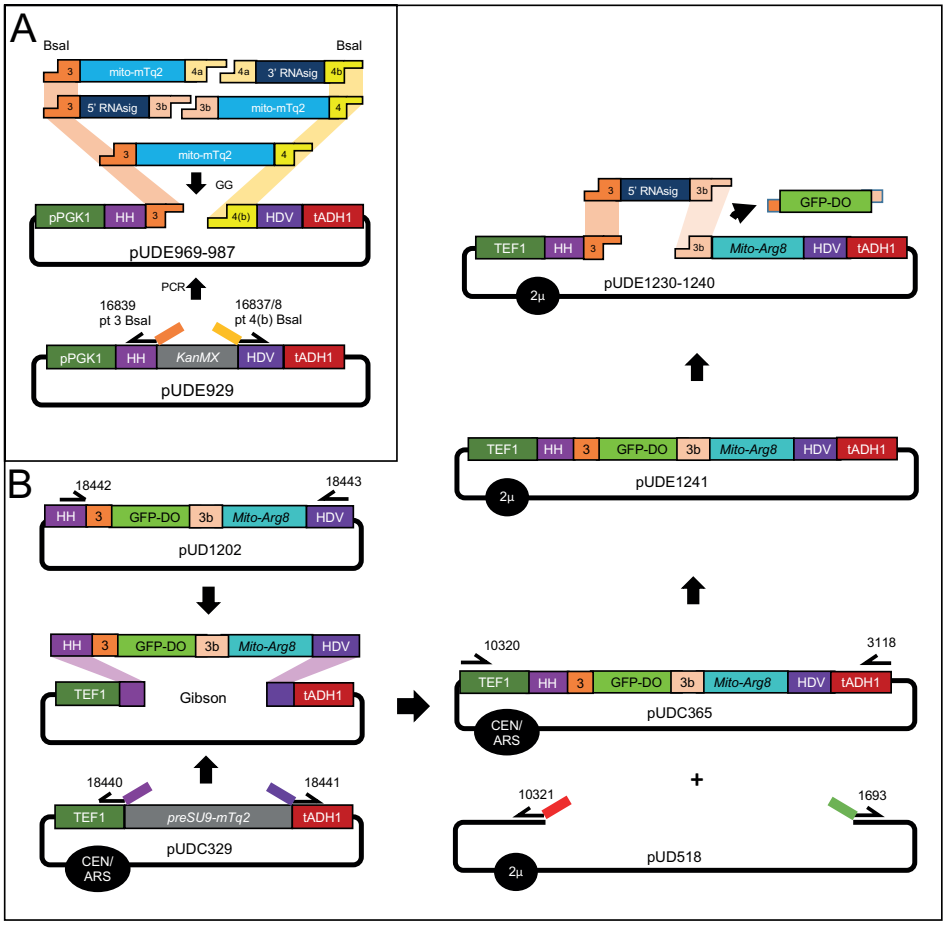


Figure S5.1.Construction strategy of libraries of mito-mTq2 mRNA (A) and mito-*ARG8* mRNA (B) with different RNA import signals.

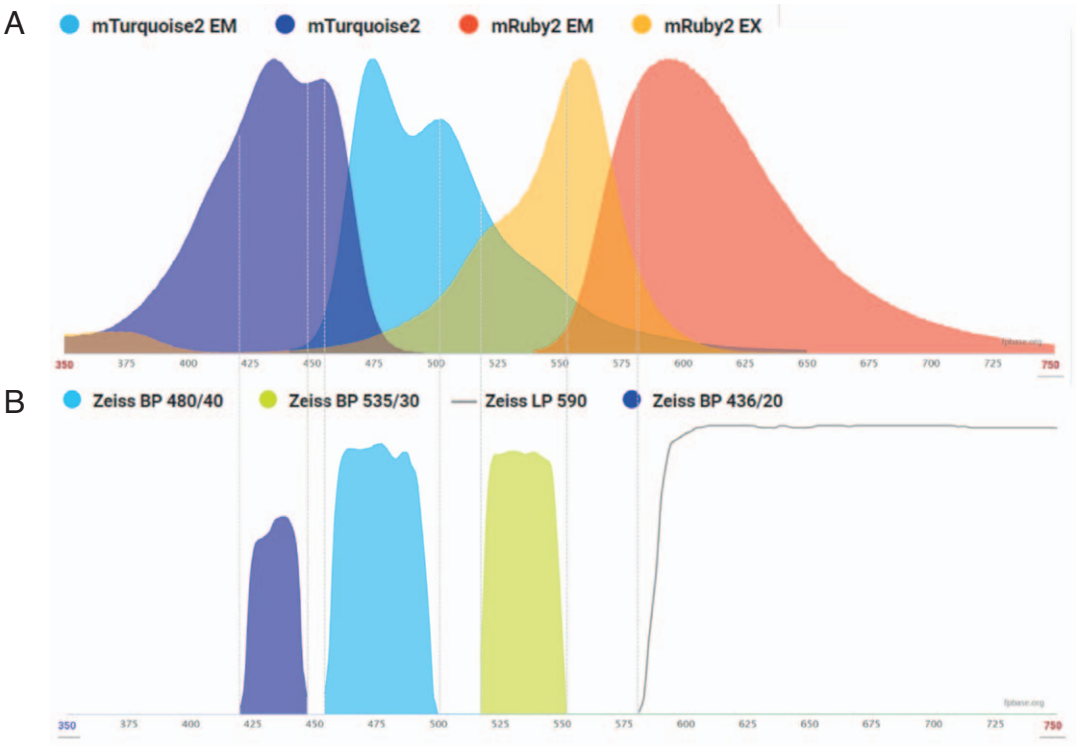


Figure S5.2. Spectral overview of (A) fluorophores and (B) filters used for *in vivo* co-localization assays. The x-axes display wavelength in nanometers, the spectral plot is generated using FPBase.

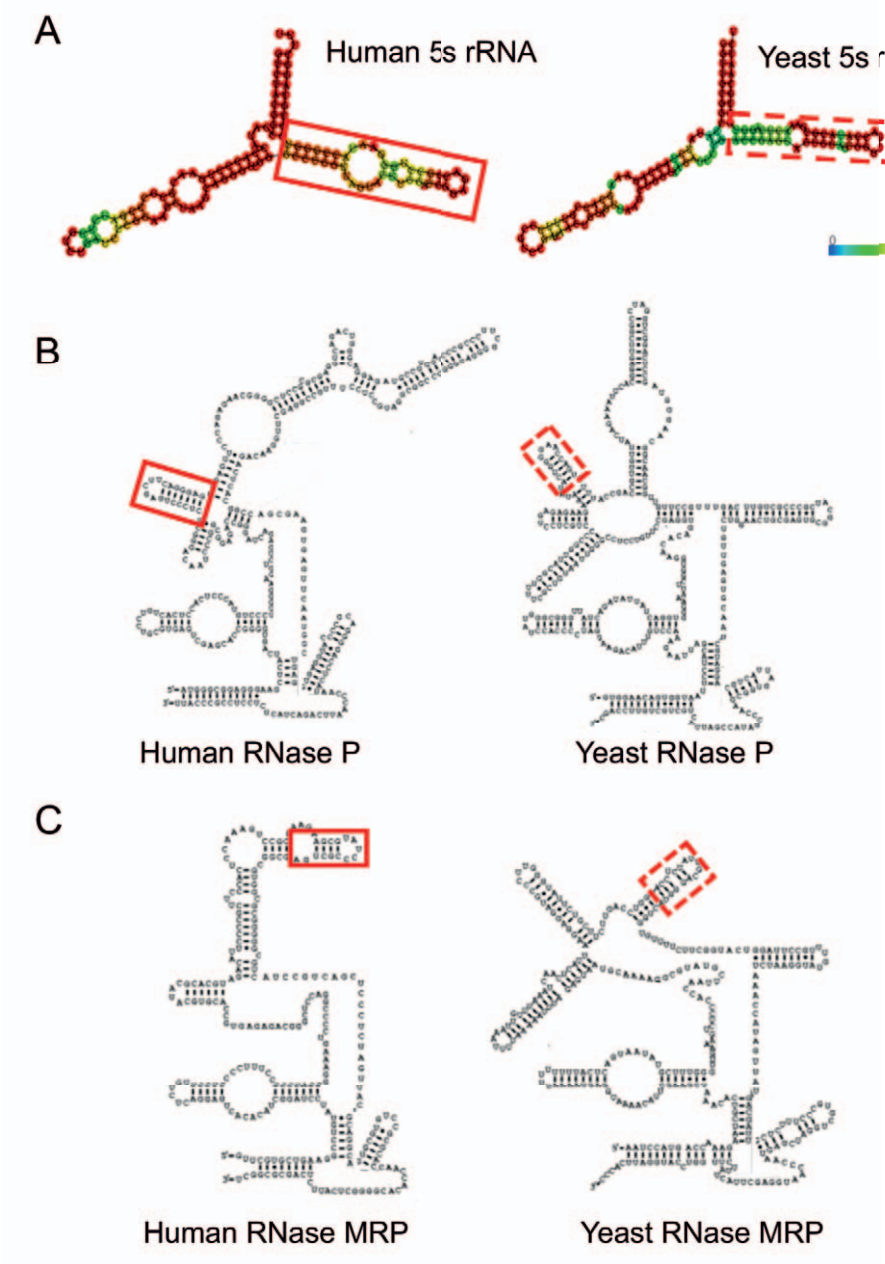


Figure S5.3. Structural comparison between human and *S. cerevisiae* putative RNA import signals. A) Human 5S rRNA (accession no. NR_023363) and *S. cerevisiae* 5S rRNA (accession no. NR_132215.1). The putative h5S rRNA MAM as reported Zelenka *et al.* [26] is displayed in a solid red box. In order to extrapolate the h5S rRNA stem to a yeast equivalent, a comparative analysis was done between the full 5S rDNA sequences as found in and in yeast. Global alignment of these two exactly 121 nt sequences scored 75% similarity. A structural comparison of these native 5S rRNA components revealed similar structures as well. In h5S rRNA the sequence of the putative RNA signal mapped to a highly similar secondary structure found in yeast 5S rRNA, displayed in a dashed red box. Base-pairing probabilities are denoted according to the color scale as indicated. Structural analysis was done using RNAfold (Vienna RNA secondary structure server [118, 119]). B,C) Structural comparison between RNA components of human RNase P and *S. cerevisiae* RNase P (B) and human RNase MRP and *S. cerevisiae* RNase MRP (C). Putative hRP and hMRP stems as reported by Wang *et al.* (2010) [24] are indicated with a solid red box. Subsequent mapping of this substructure to the yeast RNA variants allowed for extrapolation of the putative signals. Apart from structural similarity, information on identified conserved intramolecular RNA domains justify the choice of specific regions as extrapolated signals to be examined for ability to stimulate RNA transport to the mitochondria. Based on structural similarity, and identified conserved intramolecular domains, the putative yeast equivalents are displayed in a dashed red box. Adapted from Esakova and Krasilnikov [120].

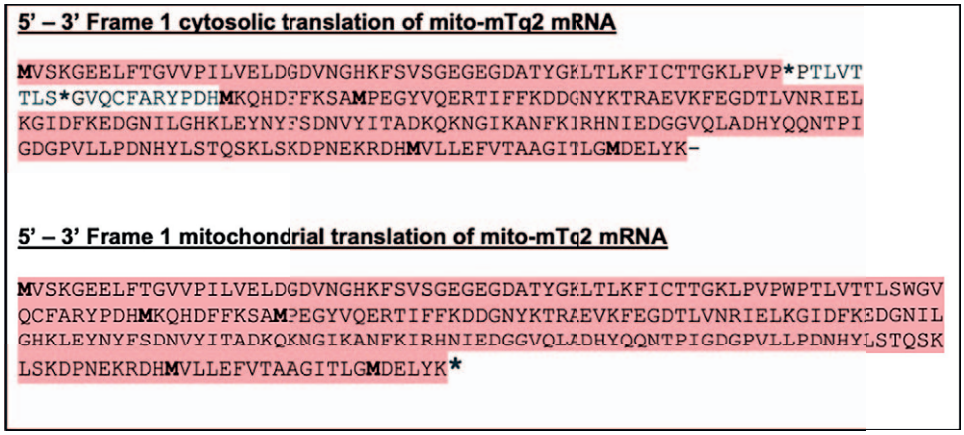


Figure S5.4. Translation frames of mito-mTq2 mRNA. Top: 5' – 3' Frame 1 translation of mito-mTq2 using yeast cytosolic genetic codon usage. Bottom: 5' – 3' Frame 1 translation of mito-mTq2 using yeast mitochondrial genetic codon usage. Open reading frames are indicated in red, potential START codons are indicated with a bold-face M, STOP codons are indicated with an asterisk (*).

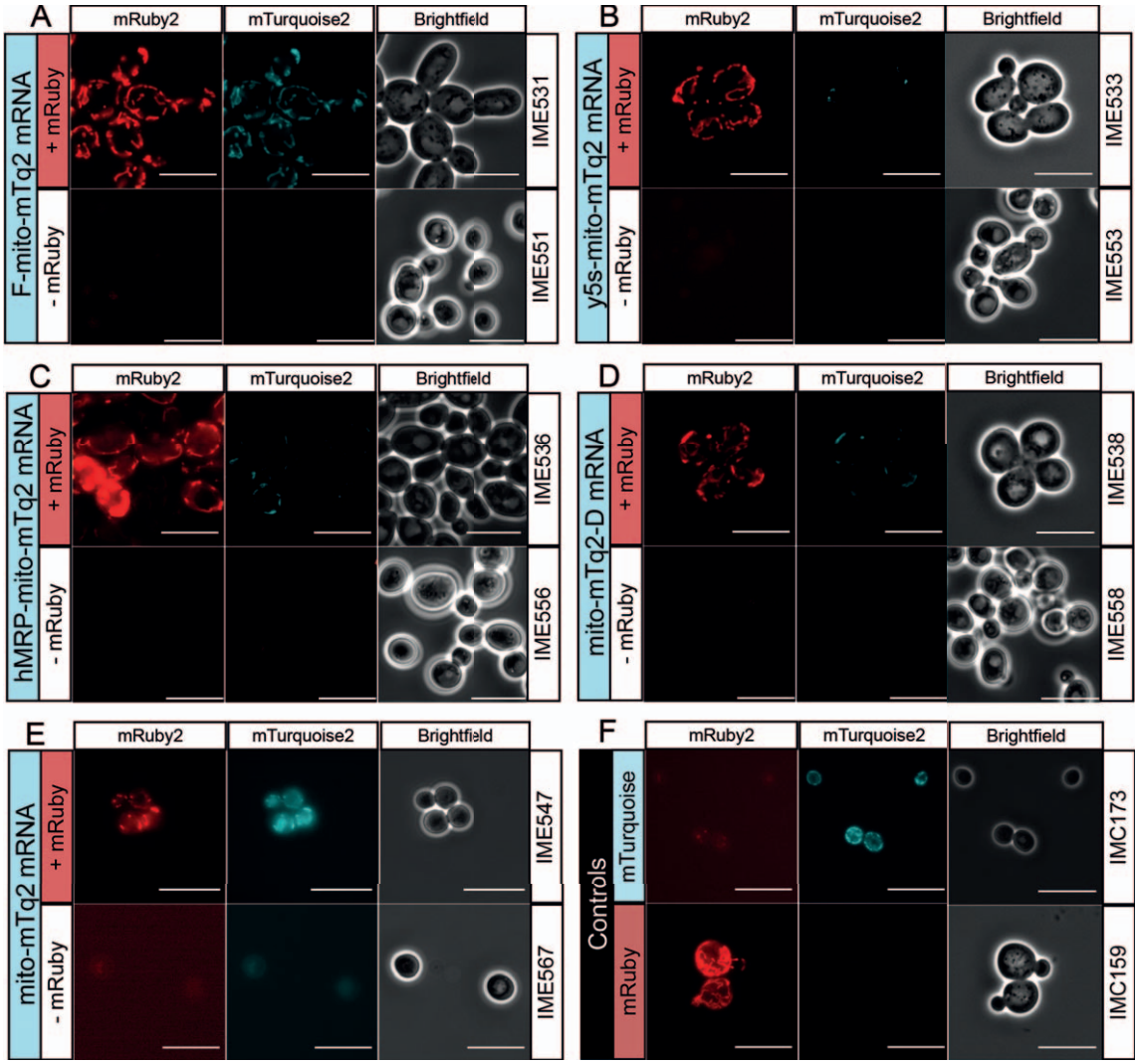
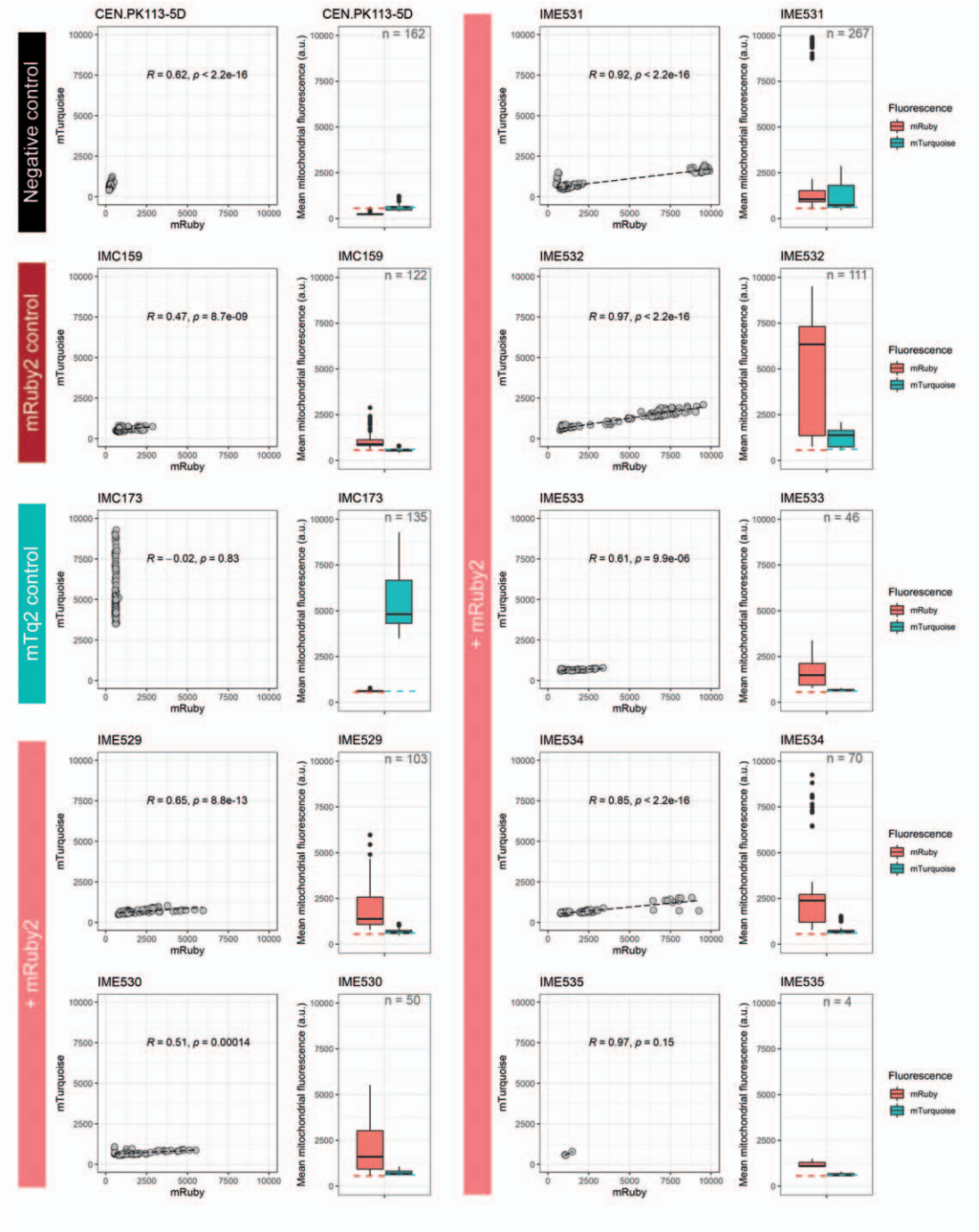
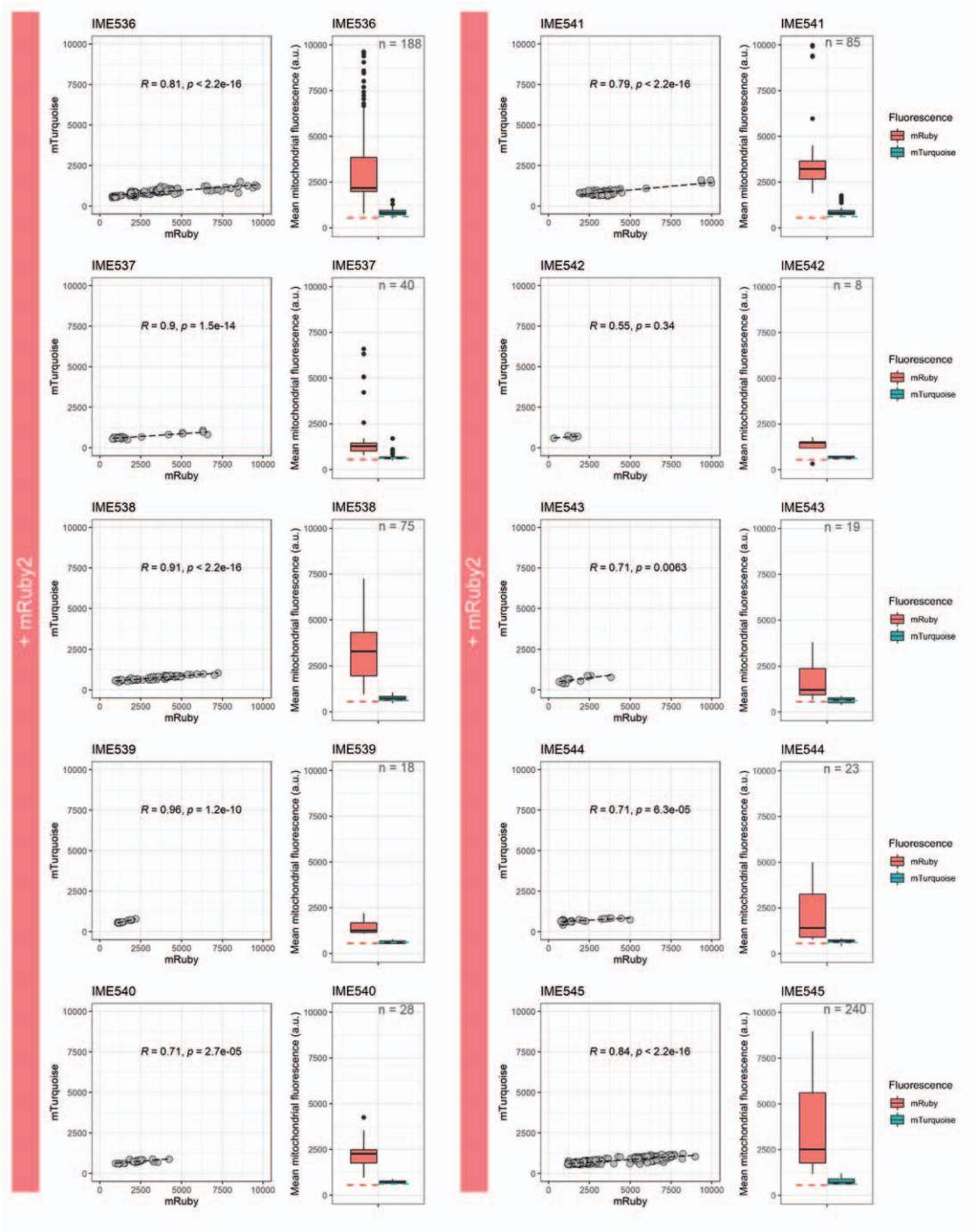


Figure S5.5. Fluorescence microscopy images (1000x) in red (mRuby) and blue (mTurquoise) channels and brightfield images of strains expressing mito-mTq2 fused to different mitochondrial RNA import signals. The fluorescence of strains expressing each signal was measured either in the presence (IME531-IME547, panels A to E, "+mRuby") or absence (IME551-IME567, panels A to E "-mRuby") of mRuby2, along with a positive control targeting the preCOX4-mTq2 protein to the mitochondria (IMC173, panel F, mTurquoise), a negative control omitting the mito-mTq2 RNA (IMC159, panel F, mRuby). All pictures were obtained with the same settings and an exposure times of 1200 ms for mTq2 and 500 ms for mRuby2. Images were processed using the same settings for brightness and contrast, where any pixel values of lower than 1000 were omitted to remove background fluorescence. Scale bars represent 10 µm.



(Figure continues on next page)



(Figure continues on next page)

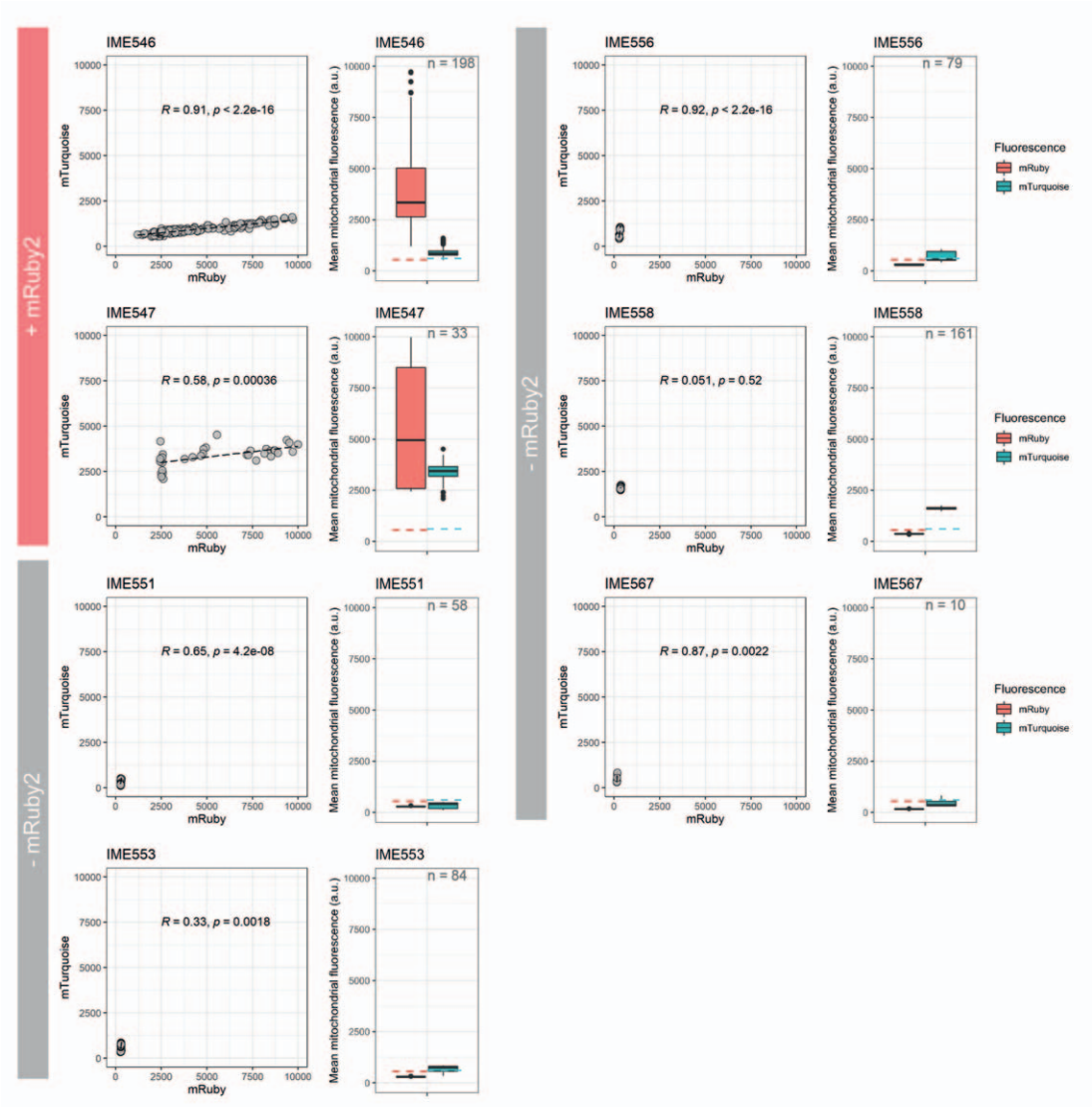
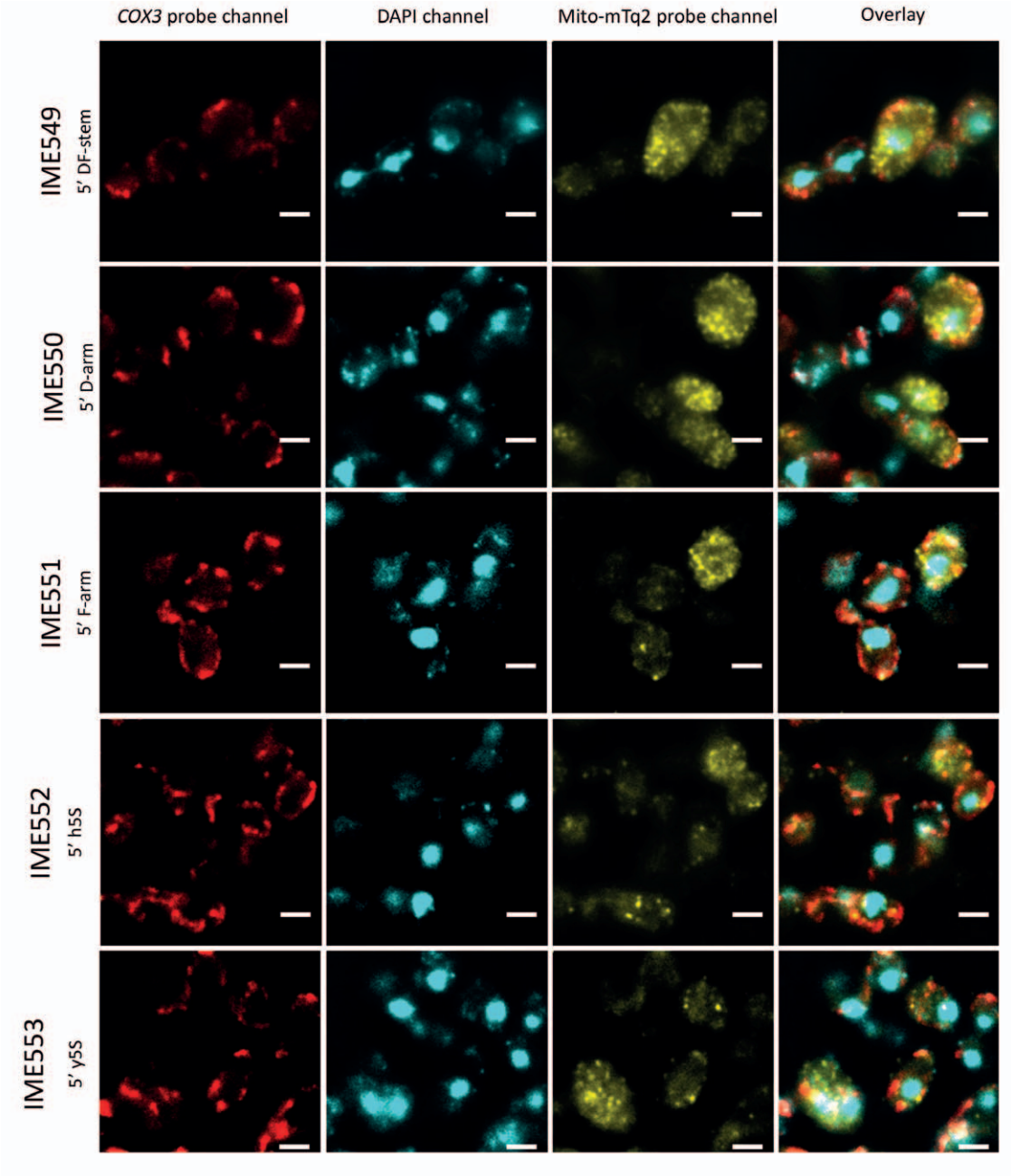
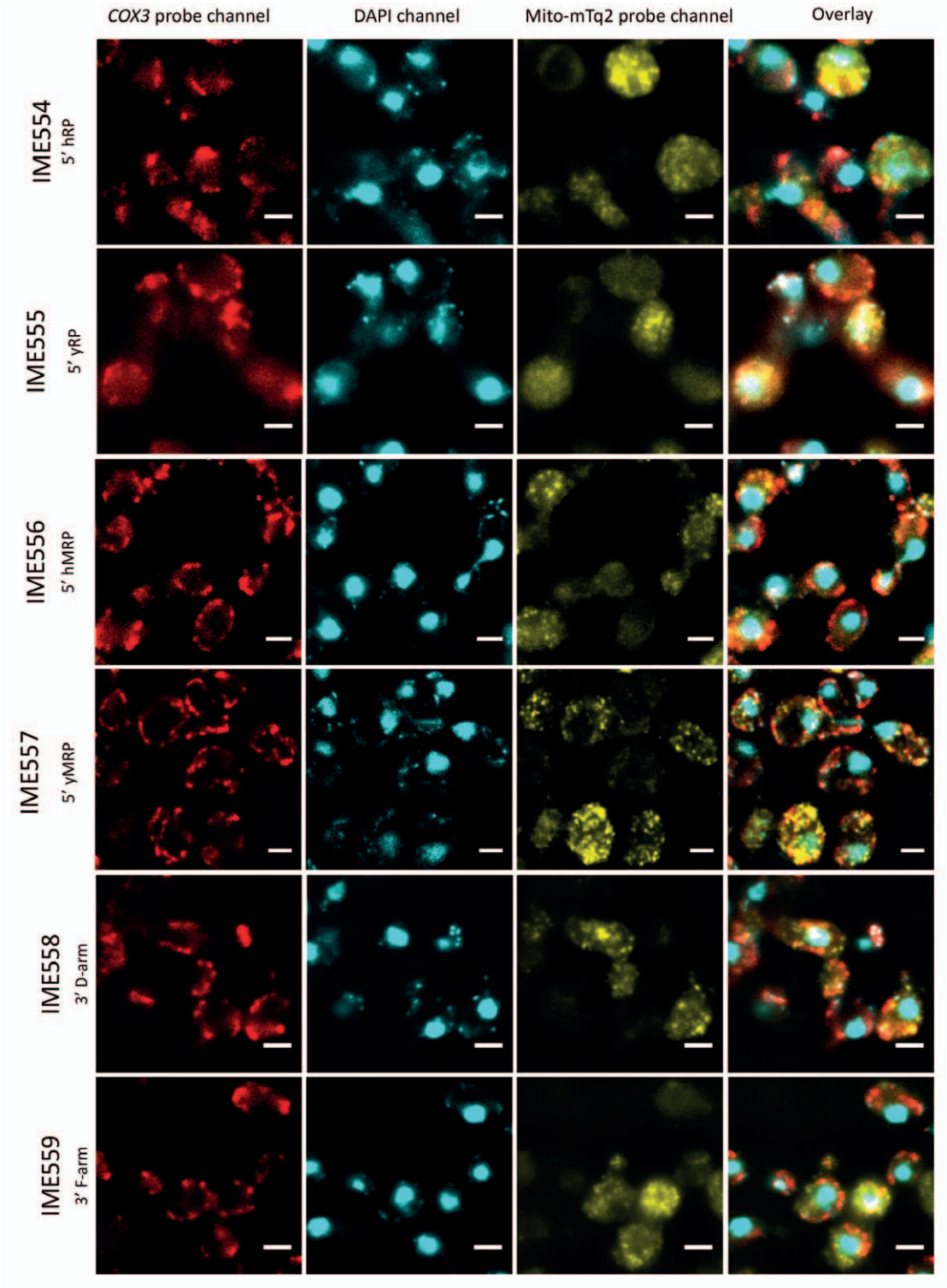


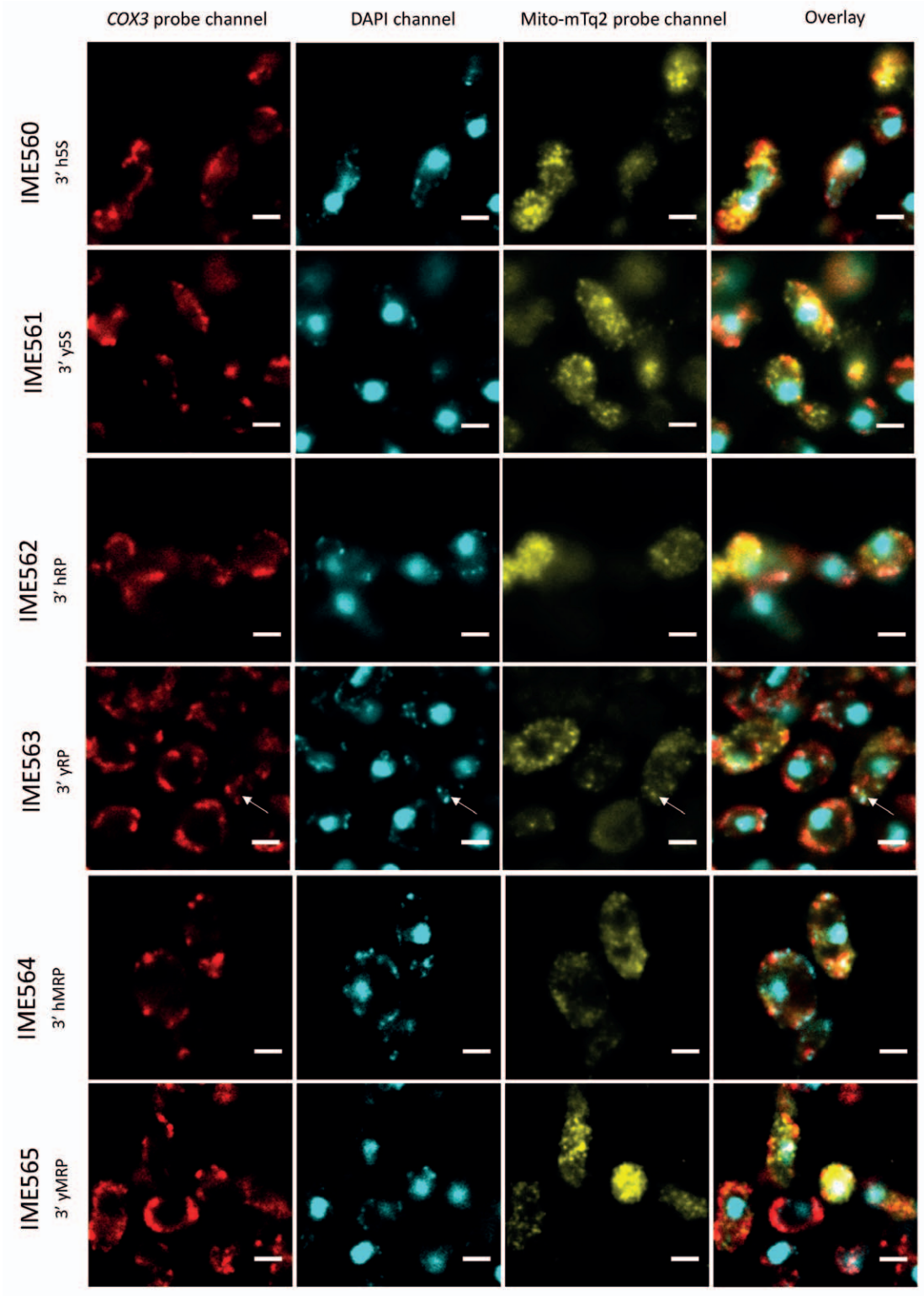
Figure S5.6. Distribution of the mean mitochondrial fluorescence in the different strains in the mRuby2 (red) and mTurquoise2 (blue) channels, determined from microscopic analysis and measuring mean pixel intensities in the red and blue channels. Left: Dot plots showing the correlation of the mean measured pixel intensity in the red channel (mRuby) versus the measured mean pixel intensity in the blue (mTurquoise2) channel. A linear trendline is shown as dashed line, R represents the Pearson correlation coefficient, p represents the p-value of the calculated correlation, a $p < 0.05$ is deemed significant. One dot represents the mitochondrial content of one cell. Right: The distribution of the mean fluorescence intensity of mRuby (red) and mTurquoise (blue) presented as boxplots, where the line represents the median value, and the edges of the boxes represent the 25th and 75th percentile. Outliers are shown as dots. The dashed line represents the median fluorescence value of the negative controls as a reference (IMC173 for Ruby fluorescence, IMC159 for mTq2 fluorescence). The strains presented contain mito-mTq2 fused to the same RNA targeting signal either in the presence (IME529-IME547) or absence (IME551-IME567) of mRuby2.



(Figure continues on next page)



(Figure continues on next page)



(Figure continues on next page)

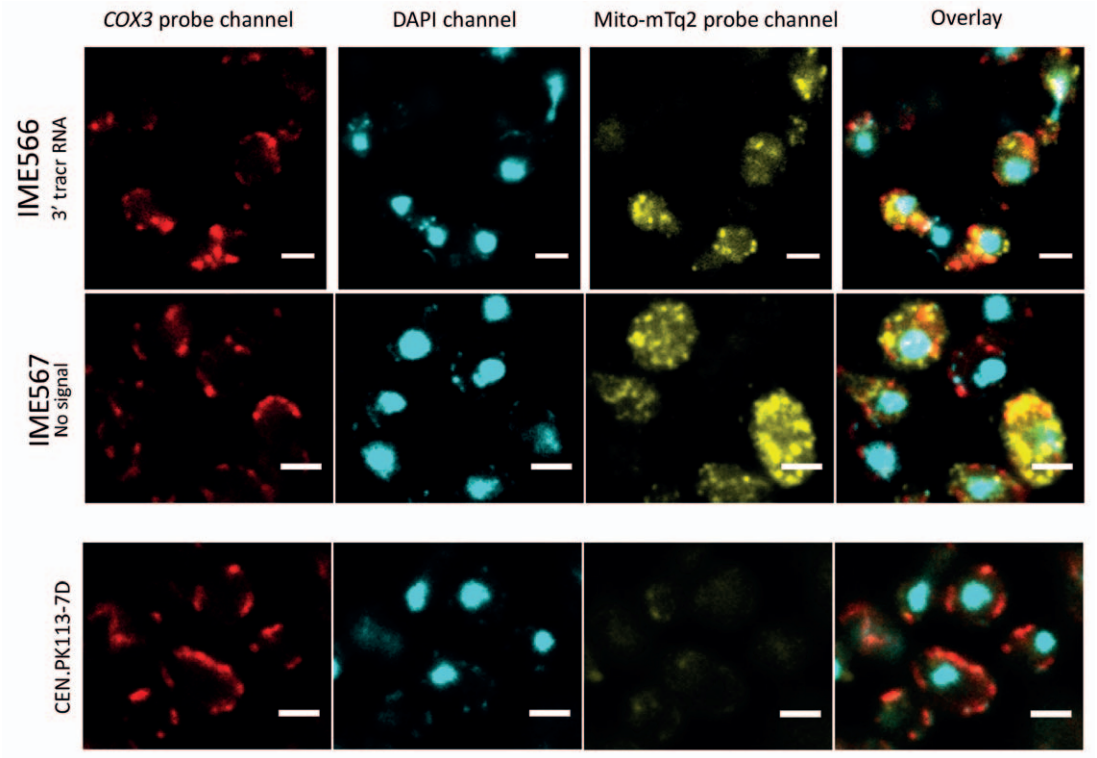


Figure S5.7. FISH analysis of strains expressing mito-mTq2 mRNA with different RNA signals and negative control CEN.PK113-7D, omitting mito-mTq2 mRNA. Acquisition in the *COX3* probe channel, mito-mTq2 probe and the DAPI channel were respectively done with exposure times of 1000, 600 and 100 seconds, at a magnification of 100x. Every image was derived from a single slice in a Z-stack after iterative deconvolution. The white scale bar indicates 5 µm.

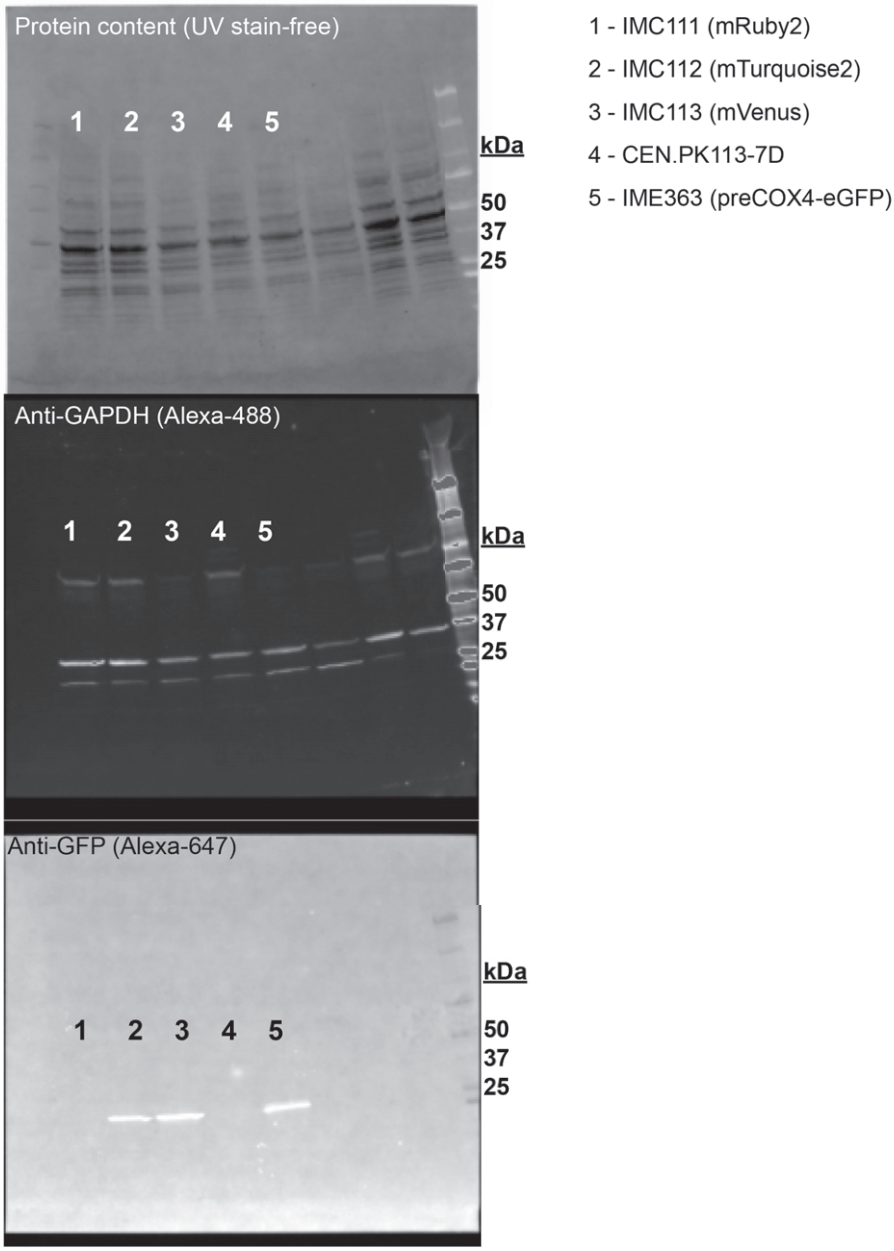
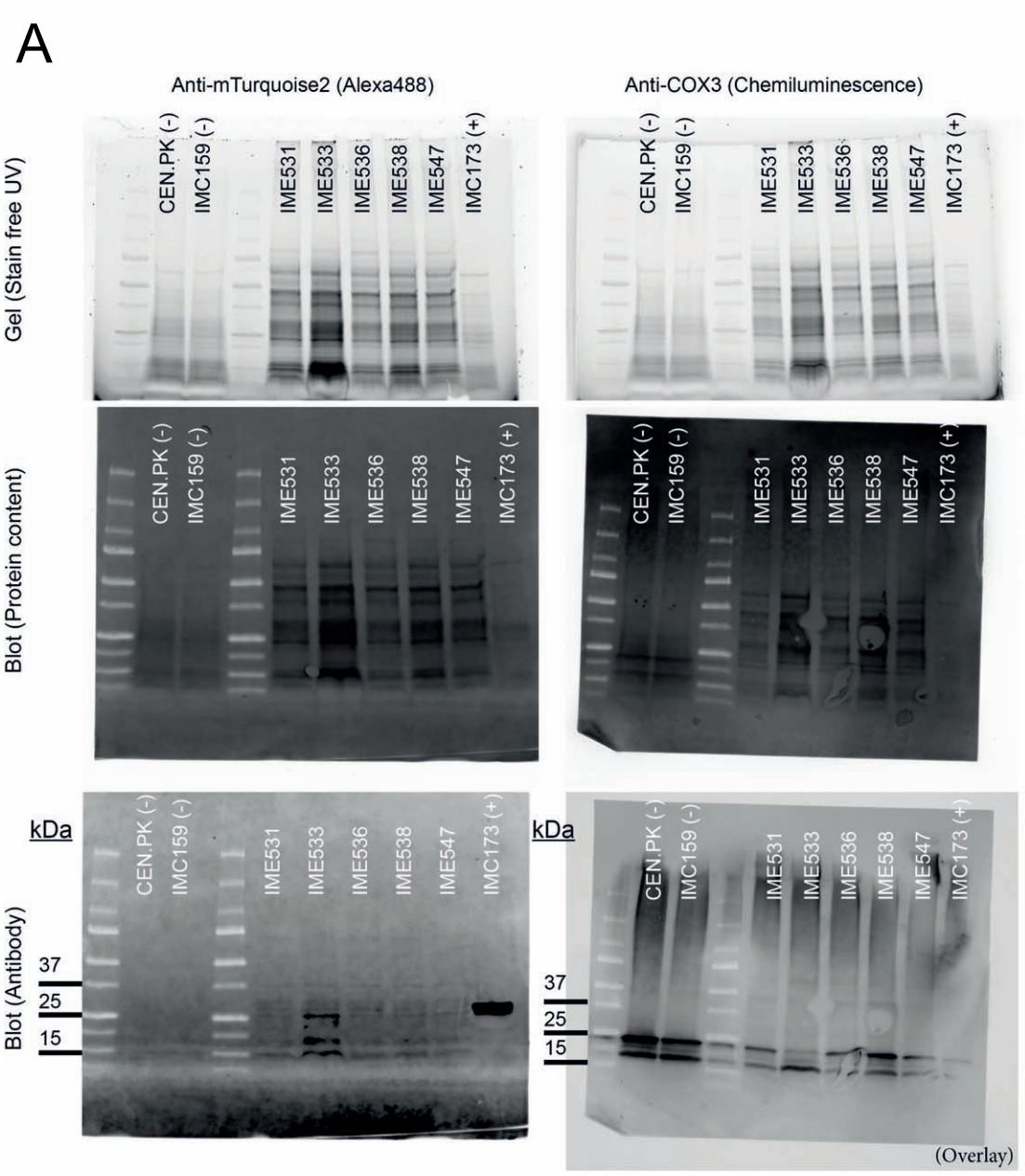


Figure S5.8. Western blot of fluorescent proteins using anti-GFP. Positive control strain IME363 expressing eGFP (26.9 kDa) was compared to control strains IMC112 and IMC113 resp. expressing mTq2 (26.9 kDa) and Venus (26.8 kDa) for capability of anti-GFP to recognize the GFP-derivatives. Negative control strain IMC111 expressing mRuby2 (26.5 kDa) was not recognized by anti-GFP (bottom). 25 µg protein was loaded in each lane at the start of PAGE (top), anti-GAPDH (middle) was used as a protein loading control (MW_{GAPDH}= 36 kDa). The reference ladder displays protein size in kDa.



(Figure continues on next page)

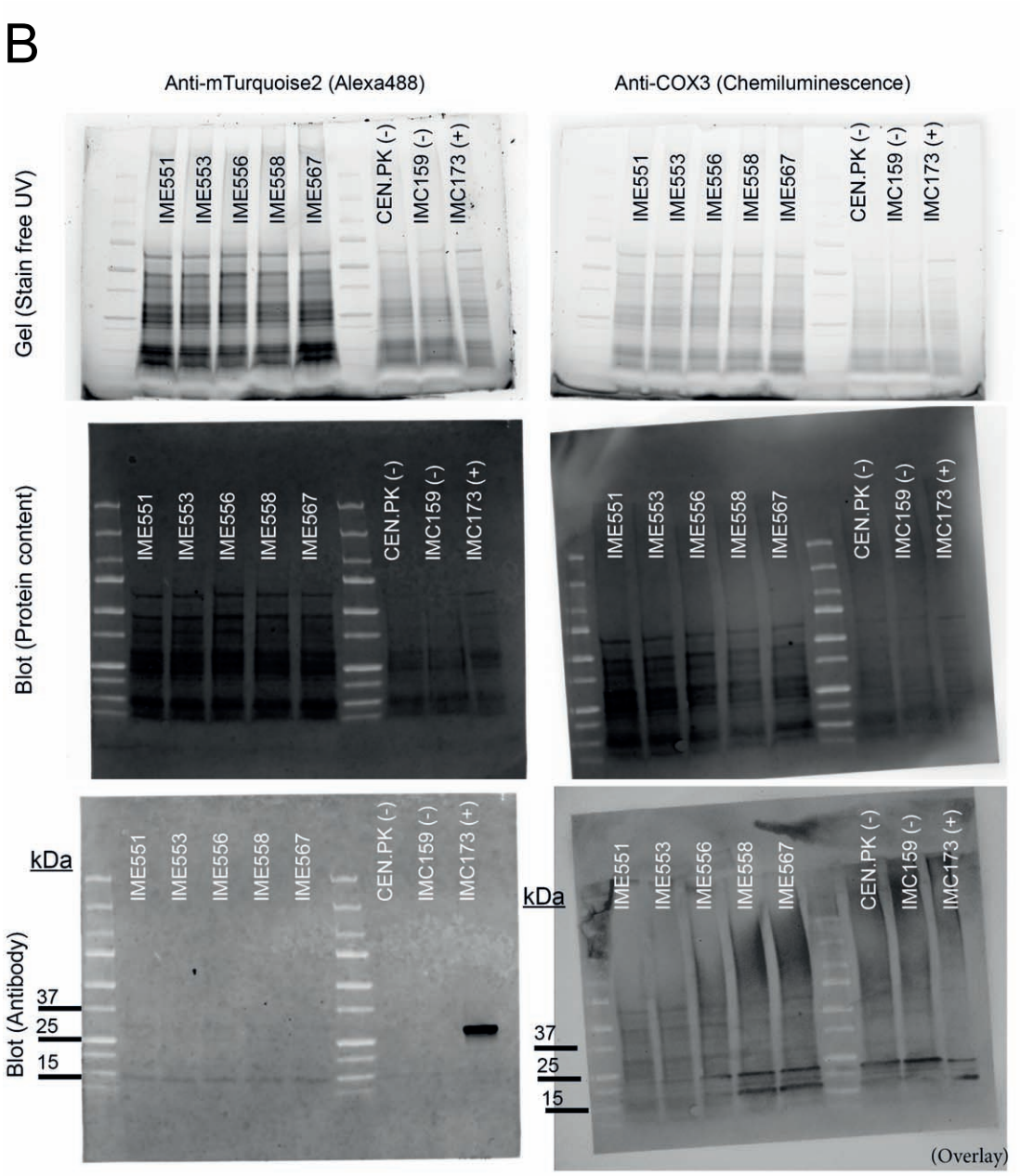


Figure S5.9 Western blot analysis of mTurquoise2 expression in strains expressing mito-mTq2 mRNA in with the presence (panel A) or absence (panel B) of the preCOX4-mRuby2 protein. Negative controls (-) include CEN.PK113-7D (no fluorescent protein) and IMC159 (preCOX4-mRuby2 only), and the positive control IMC173 indicated with (+) expresses preSU9-mTq2. Top images: full protein content of isolated mitochondria separated on an SDS-PAGE gel. Middle images: full protein content of isolated mitochondria transferred on a blotting membrane. Bottom images: membranes blotted with antibody and imaged with the corresponding settings. The left column shows isolated mitochondria blotted with an anti-GFP antibody conjugated with an Alexa-488 probe, imaged at 488 nm, the right column shows blotted with an anti-Cox3p (subunit 3 of the cytochrome *c* oxidase of the respiratory chain) antibody-HRP complex, treated with substrate and imaged using chemiluminescence as a loading control. The chemiluminescence image was overlaid with the blot to allow for visibility of the ladder. Approximate protein sizes of the observed bands in kDa are indicated on the left, the used ladder was the Precision Plus Protein Dual Color Standards (Bio-Rad). Each lane was loaded with 25 µg protein, except for IMC173, that was loaded with 10 µg protein. Expected protein sizes are 26.9 kDa for mTq2 and 30 kDa for Cox3p.

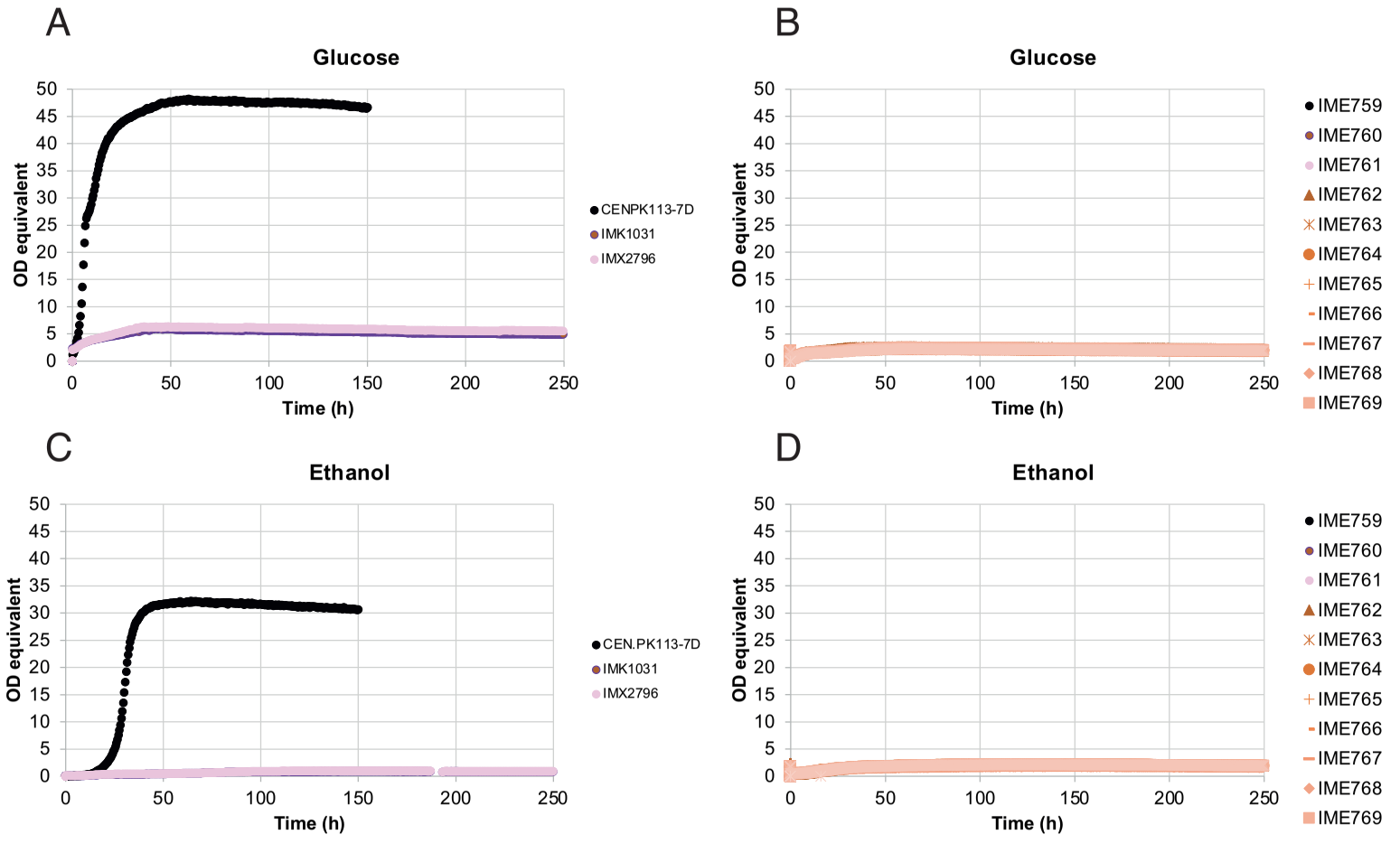


Figure S5.10 Growth of strains with mito-ARG8 mRNA targeted to mitochondria, on synthetic medium without arginine in microtiterplates. A and B, controls strains CEN.PK113-7D (prototrophic reference) and the arginine auxotrophs IMK1031 *(Δarg8 Δcar2)* and IMX2796 *(Δarg8 Δcar2 YPRCtau3::preCOX4-mRuby2)* grown with glucose (A) or ethanol (C) as sole carbon source. C and D, strains IME759-IME769 which are IMX2796 transformed with different plasmids expressing mito-*ARG8* mRNA with different targeting signals, grown on glucose (B) or ethanol (D)

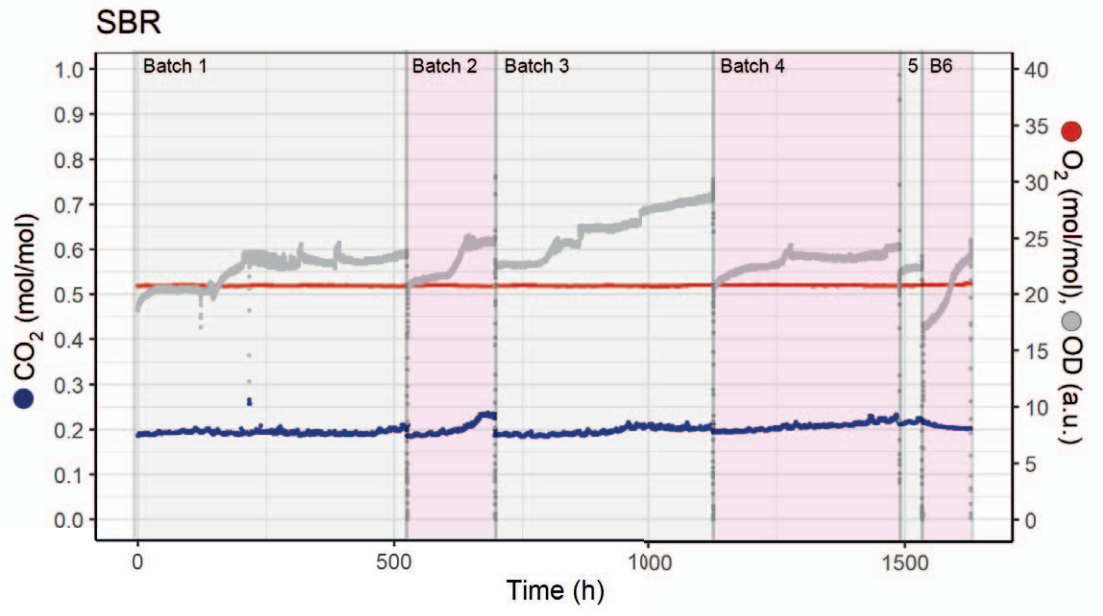


Figure S5.11. Gas and biomass (OD) profiles of the evolved population from bioreactor B grown in SBR in the absence of arginine. The population was fed with synthetic medium containing ammonium sulfate as sole nitrogen source and ethanol as sole carbon source. CO_2 and O_2 were measured continuously in the off-gas of the bioreactor, optical density was continuously measured by a non-invasive optical biomass sensor. Six successive empty-refill batch cycles are represented in alternate pink and light grey background.

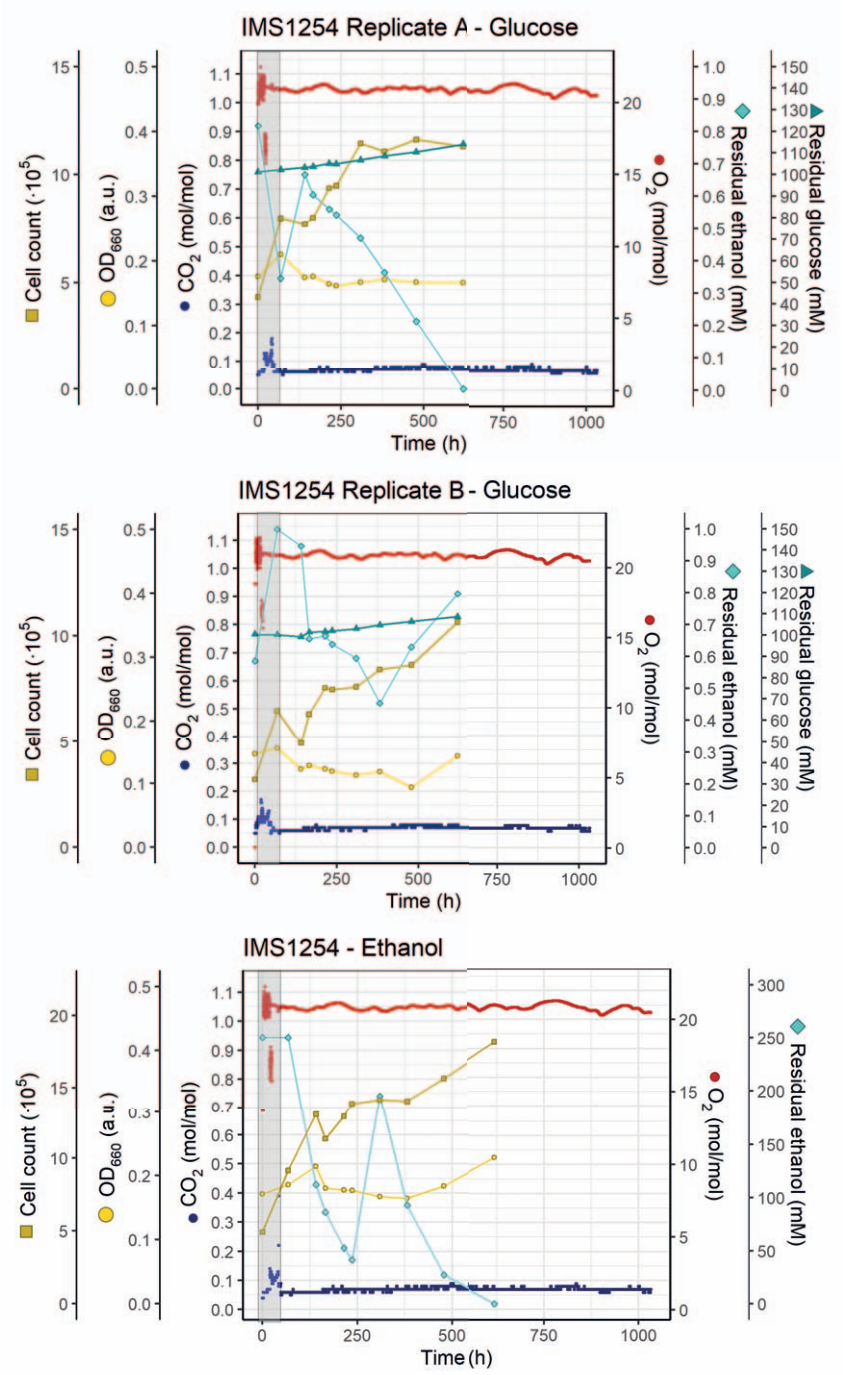


Figure S5.12. Physiological characterization of the evolved single colony isolate IMS1254 in batch cultivation in a bioreactor in the absence of arginine. Growth in synthetic medium supplied with ammonium as sole nitrogen source and either glucose (A and B) or ethanol (C) as sole carbon source. CO_2 and O_2 concentrations were measured continuously in the off-gas of the bioreactor. Optical density (OD_{660}), cell count and the residual ethanol and glucose concentrations were measured offline. The pre-cultures were grown in shake flasks in the presence of arginine. After inoculation of the bioreactor with this pre-culture, a first batch was run to deplete the residual arginine carried over from the pre-culture (grey area). At the end of this first batch, the biomass was let to settle by stopping stirring and aeration and the bioreactor was partially emptied and refilled with fresh, arginine-free medium. The white background reflects growth and metabolic activity of IMS1254 in the complete absence of arginine.

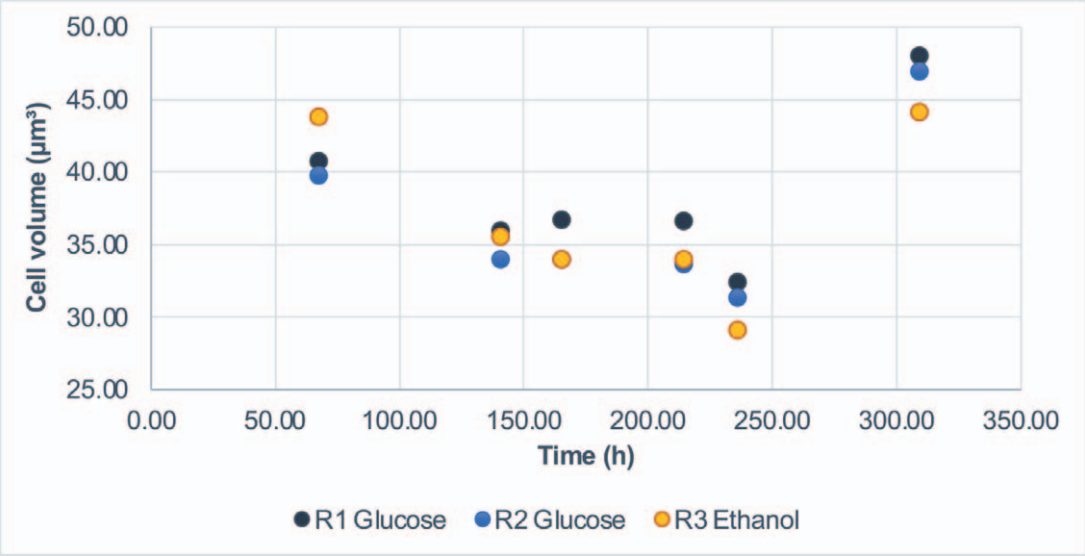


Figure S5.13. Average cell size of strain IMS1254 (bioreactor B evolved single colony isolate) in batch cultivation in a bioreactor on synthetic medium supplied with either glucose or ethanol, in the absence of arginine at different time points.

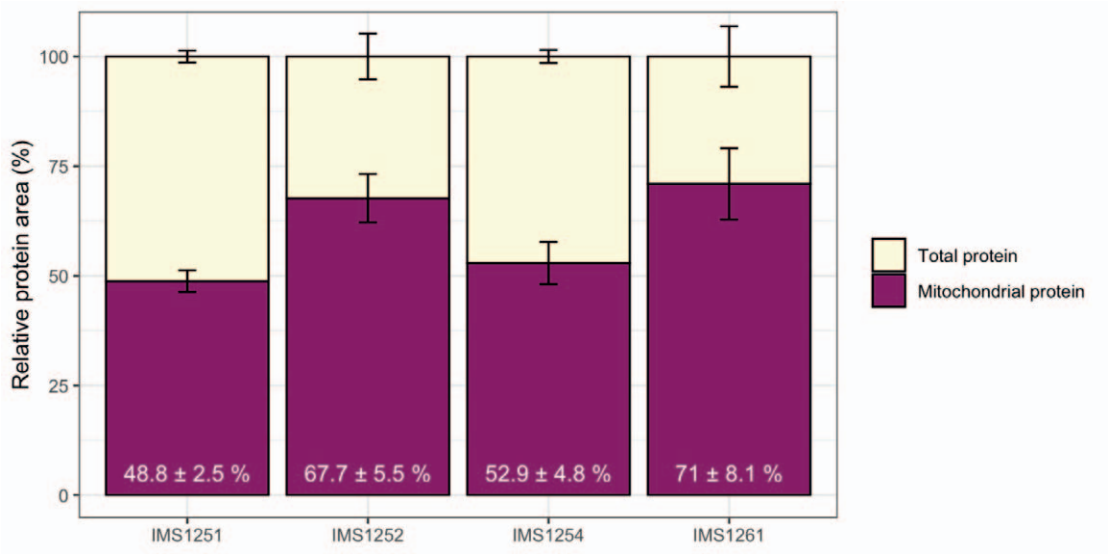


Figure S5.14. Fraction of mitochondrial protein detected in each proteome, based on the accumulative protein areas of protein that were "mitochondrial" according to their Gene Ontology descriptor versus the total protein areas of each proteome. Error bars were calculated from values obtained for biological duplicates.

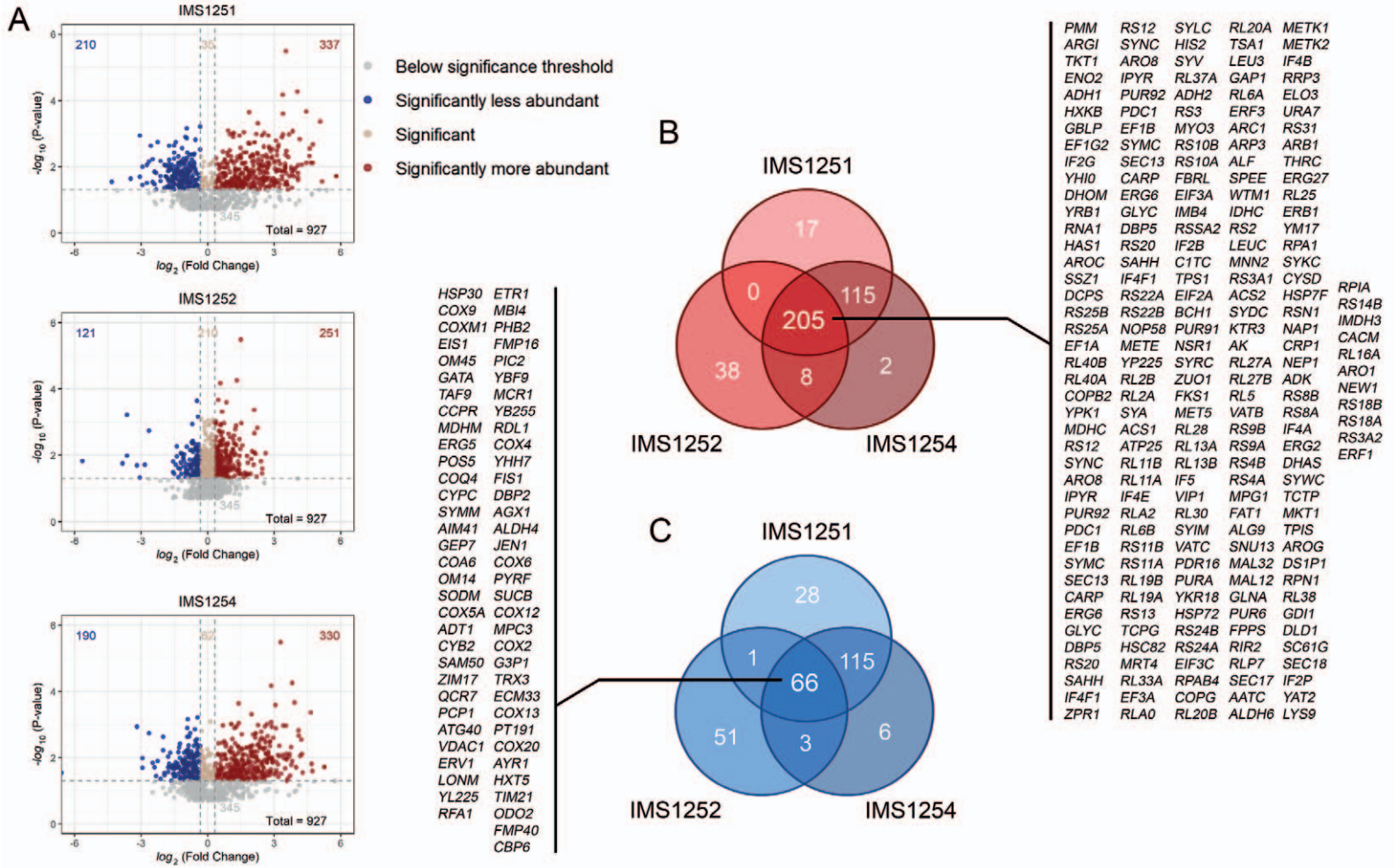


Figure S5.15. Global mitochondrial proteome changes between the evolved strains IMS1251, IMS1252 and IMS1254. A) The fold changes were normalized to the proteome of the unevolved starting population IMS1261. The *log2* of the abundance fold change between the two conditions was plotted against the significance ($-log_{10}$(p)), using a p-value threshold of < 0.05 and a fold change threshold of >1.25 (which corresponds to a log_2 fold change threshold +/- 0.32). Red depicts proteins with higher abundance in the evolved strains as compared to the starting population, blue a lower abundance, and beige a similar abundance. B) Venn diagram of the significantly more abundant proteins of each strain (indicated in red in the volcano plots of (A)). Numbers indicate the number of proteins more abundant in one, two or all three strains. The 205 genes more abundant in all three strains are listed. C) Venn diagram of the significantly less abundant proteins of each strain (indicated in blue in the volcano plots of (A)). Numbers indicate the number of proteins more abundant in one, two or all three strains. The 66 genes less abundant in all three strains are listed.

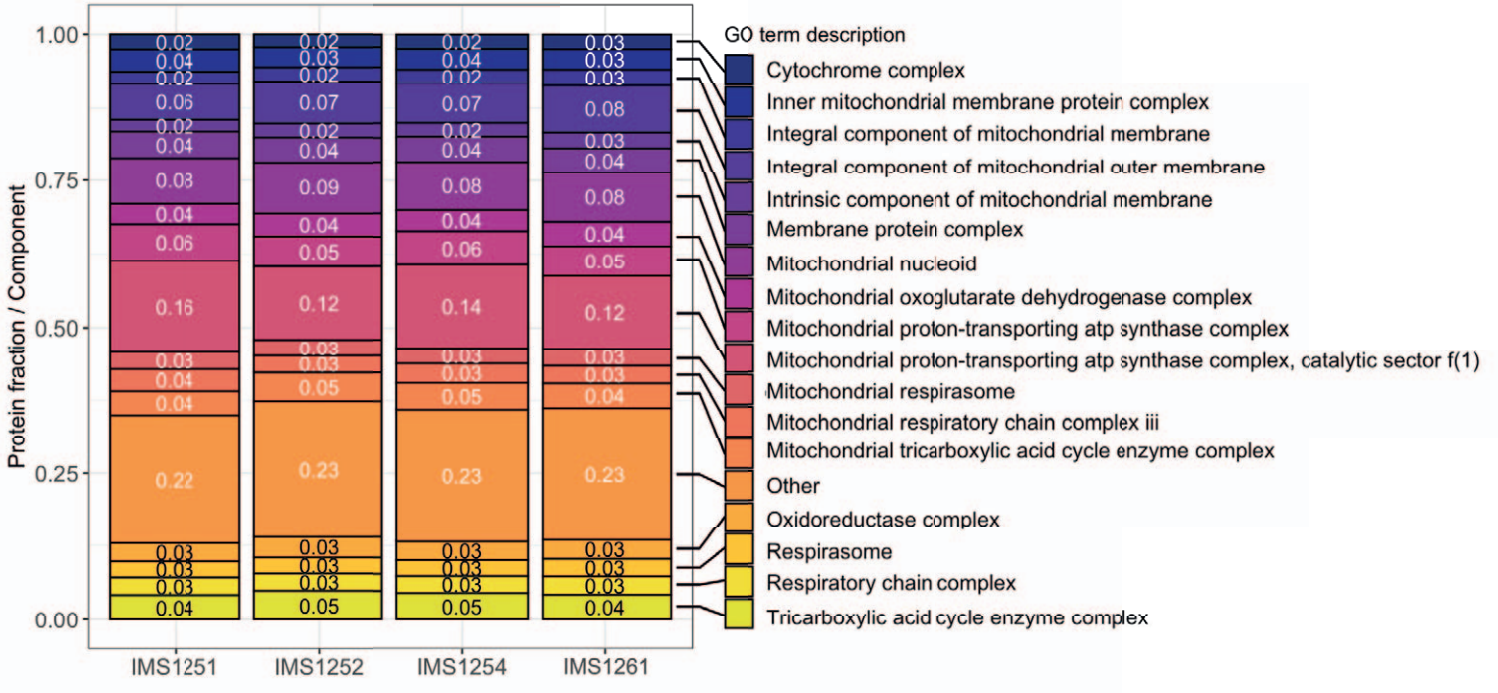


Figure S5.16. Proteome allocation within the mitochondria according to Gene Ontology (GO) terms of the type "component". The proteome allocation was calculated by adding the areas of all detected mitochondrial proteins associated to each GO-term. The category "Other" accumulates all GO-terms that had less than 2 % of the proteins associated to each term. The full proteome data are available in SI file 2.

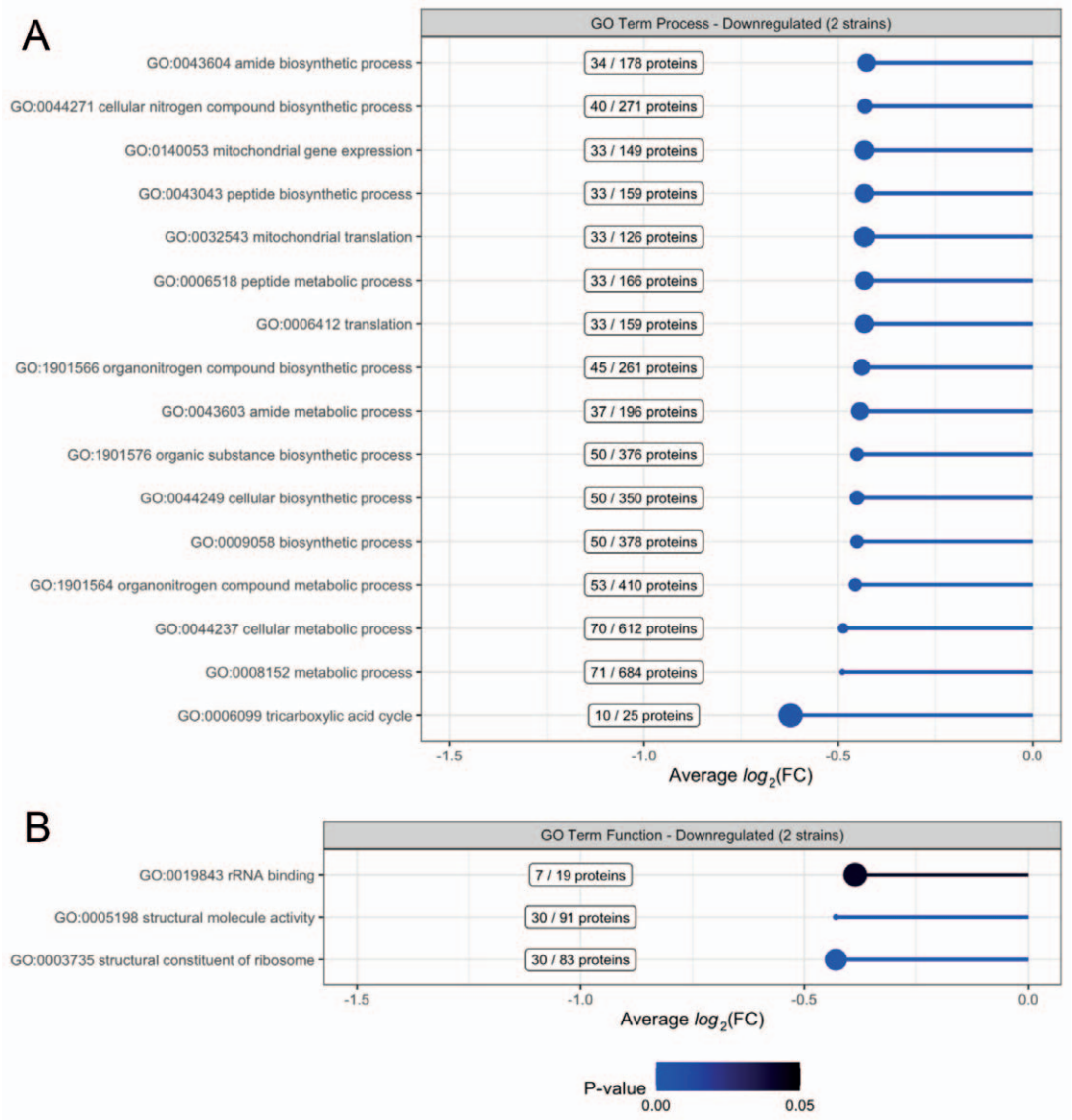


Figure S5.17. GO-term analysis of 78 proteins that were less abundant in both strains IMS1251 and IMS1254, compared to unevolved parental population IMS1261. Genes were classified according to Gene Ontology (GO) terms of the type (A) "Process" and (B) "Function". The length of each bar indicates the average measured log_2(fold change (FC)) of all genes associated with the GO-term. The size of the dot indicates the enrichment strength (number of detected proteins in each term/number of proteins of each term in the reference list), which is also indicated next to each bar. The color of each bar indicates the significance of the enrichment, all were below p = 0.05 (Bonferroni corrected). Genes associated with each GO-term are listed in SI file 2

Table S5.1.Plasmids used in this study. "YTK type #" refers to yeast Toolkit parts as described by Lee *et al.* [47]

Name	Relevant characteristics	Source
pUD518	GFP-dropout, *KanR, ColE1*	[52]
pUD1094	mitochondrial codon optimized mTq2 (mito-mTq2 mRNA), *bla, ColE1*	Synthetic
pUD1202	*HH-GFPdropout-Cox2UTL-mitoARG8-dodecamer-HDV, bla, ColE1*	Synthetic
pYTK001	*CamR, ColE1*	[47]
pGGkp265	mito-mTq2 mRNA (YTK type 3), *CamR, ColE1*	This study
pGGkp267	DF-stem import signal (YTK type 3a), *CamR, ColE1*	This study
pGGkp268	D-arm signal (YTK type 3a), *CamR, ColE1*	This study
pGGkp269	F-arm signal (YTK type 3a), *CamR, ColE1*	This study
pGGkp270	h5S rRNA signal (YTK type 3a), *CamR, ColE1*	This study
pGGkp271	y5S rRNA signal (YTK type 3a), *CamR, ColE1*	This study
pGGkp272	hRP-stem signal (YTK type 3a), *CamR, ColE1*	This study
pGGkp273	Putative yRP signal (YTK type 3a), *CamR, ColE1*	This study
pGGkp274	hMRP RP-stem signal (YTK type 3a), *CamR, ColE1*	This study
pGGkp275	Putative yMRP signal (YTK type 3a), *CamR, ColE1*	This study
pGGkp280	Mito-mTq2 mRNA ((YTK type 3b), *CamR, ColE1*	This study
pGGkp283	D-arm signal (YTK type 4a), *CamR, ColE1*	This study
pGGkp284	F-arm signal (YTK type 4a), *CamR, ColE1*	This study
pGGkp285	h5S rRNA signal (YTK type 4a), *CamR, ColE1*	This study
pGGkp286	y5S rRNA signal (YTK type 4a), *CamR, ColE1*	This study
pGGkp287	hRP-stem signal (YTK type 4a), *CamR, ColE1*	This study
pGGkp288	Putative yRP signal (YTK type 4a), *CamR, ColE1*	This study
pGGkp289	hMRP RP-stem signal (YTK type 4a), *CamR, ColE1*	This study
pGGkp290	Putative yMRP signal (YTK type 4a), *CamR, ColE1*	This study
pGGkp291	tracrRNA signal (YTK type 4a), *CamR, ColE1*	This study
pUDC191	*pCCW12-mRuby2-tENO1 URA3, CEN6/ARS4, bla, ColE1*	[43]
pUDC192	*pTEF2-mTurquoise2-tSSA1 URA3, CEN6/ARS4, bla, ColE1*	[43]
pUDC193	*pTEF1-Venus-tTDH1 URA3, CEN6/ARS4, bla, ColE1*	[43]
pUDC286	*pTDH3-preCOX4-mRuby2-tADH1 URA3, CEN6/ARS4, bla, ColE1*	[44]
pUDC329	*pTEF1-preSU9-ymTq2-tENO2, URA3, CEN6/ARS4, bla, ColE1*	[45]
pUDC365	*pTEF1-HH-GFP*dropout*-Cox2UTL-mitoARG8-dodecamer-HDV-tENO2, URA3, CEN6/ARS4, KanR, ColE1*	This study
pUDC434	*pTDH3-preCOX4MTS-ARG8-tADH1, URA3, CEN6/ARS4, bla, ColE1*	[42]

pUDE929	*pPGK1-HH-dKanMX-HDV-tADH1, HygR, 2µ, bla, ColE1*	This study
pUDE969	*pPGK1-HH-mito-mTq2-HDV-tADH1, HygR, 2µ, bla, ColE1*	This study
pUDE970	*pPGK1-HH-DF-mito-mTq2-HDV-tADH1, HygR, 2µ, bla, ColE1*	This study
pUDE971	*pPGK1-HH-D-mito-mTq2-HDV-tADH1, HygR, 2µ, bla, ColE1*	This study
pUDE972	*pPGK1-HH-F-mito-mTq2-HDV-tADH1, HygR, 2µ, bla, ColE1*	This study
pUDE973	*pPGK1-HH-h5S-mito-mTq2-HDV-tADH1, HygR, 2µ, bla, ColE1*	This study
pUDE974	*pPGK1-HH-y5S-mito-mTq2-HDV-tADH1, HygR, 2µ, bla, ColE1*	This study
pUDE975	*pPGK1-HH-hRP-mito-mTq2-HDV-tADH1, HygR, 2µ, bla, ColE1*	This study
pUDE976	*pPGK1-HH-yRP-mito-mTq2-HDV-tADH1, HygR, 2µ, bla, ColE1*	This study
pUDE977	*pPGK1-HH-hMRP-mito-mTq2-HDV-tADH1, HygR, 2µ, bla, ColE1*	This study
pUDE978	*pPGK1-HH-yMRP-mito-mTq2-HDV-tADH1, HygR, 2µ, bla, ColE1*	This study
pUDE979	*pPGK1-HH-mito-mTq2-D-HDV-tADH1, HygR, 2µ, bla, ColE1*	This study
pUDE980	*pPGK1-HH-mito-mTq2-F-HDV-tADH1, HygR, 2µ, bla, ColE1*	This study
pUDE981	*pPGK1-HH-mito-mTq2-h5S-HDV-tADH1, HygR, 2µ, bla, ColE1*	This study
pUDE982	*pPGK1-HH-mito-mTq2-y5S-HDV-tADH1, HygR, 2µ, bla, ColE1*	This study
pUDE983	*pPGK1-HH-mito-mTq2-hRP-HDV-tADH1, HygR, 2µ, bla, ColE1*	This study
pUDE984	*pPGK1-HH-mito-mTq2-yRP-HDV-tADH1, HygR, 2µ, bla, ColE1*	This study
pUDE985	*pPGK1-HH-mito-mTq2-hMRP-HDV-tADH1, HygR, 2µ, bla, ColE1*	This study
pUDE986	*pPGK1-HH-mito-mTq2-yMRP-HDV-tADH1, HygR, 2µ, bla, ColE1*	This study
pUDE987	*pPGK1-HH-mito-mTq2-tracr-HDV-tADH1, HygR, 2µ, bla, ColE1*	This study
pUDE1230	*pTEF1-HH-D1F1 -Cox2UTL-mitoARG8-dodecamer-HDV-tENO2, URA3, 2µ, KanR, ColE1*	This study
pUDE1231	*pTEF1-HH-D-Cox2UTL-mitoARG8-dodecamer-HDV-tENO2, URA3, 2µ, KanR, ColE1*	This study
pUDE1232	*pTEF1-HH-F -Cox2UTL-mitoARG8-dodecamer-HDV-tENO2, URA3, 2µ, KanR, ColE1*	This study
pUDE1233	*pTEF1-HH-h5S -Cox2UTL-mitoARG8-dodecamer-HDV-tENO2, URA3, 2µ, KanR, ColE1*	This study
pUDE1234	*pTEF1-HH-y5S -Cox2UTL-mitoARG8-dodecamer-HDV-tENO2, URA3, 2µ, KanR, ColE1*	This study
pUDE1235	*pTEF1-HH-hRP -Cox2UTL-mitoARG8-dodecamer-HDV-tENO2, URA3, 2µ, KanR, ColE1*	This study
pUDE1236	*pTEF1-HH-yRP -Cox2UTL-mitoARG8-dodecamer-HDV-tENO2, URA3, 2µ, KanR, ColE1*	This study
pUDE1237	*pTEF1-HH-hMRP -Cox2UTL-mitoARG8-dodecamer-HDV-tENO2, URA3, 2µ, KanR, ColE1*	This study
pUDE1238	*pTEF1-HH-yMRP -Cox2UTL-mitoARG8-dodecamer-HDV-tENO2, URA3, 2µ, KanR, ColE1*	This study
pUDE1239	*pTEF1-HH-F1 -Cox2UTL-mitoARG8-dodecamer-HDV-tENO2, URA3, 2µ, KanR, ColE1*	This study
pUDE1240	*pTEF1-HH-tr93 -Cox2UTL-mitoARG8-dodecamer-HDV-tENO2, URA3, 2µ, KanR, ColE1*	This study
pUDE1241	*pTEF1-HH-GFPdropout-Cox2UTL-mitoARG8-dodecamer-HDV-tENO2, URA3, 2µ, KanR, ColE1*	This study
pUDR514	*2x gRNA-YPRCtau3, KanMX, 2µ, bla, ColE1*	[48]
pMEL13	*gRNA-CAN1.Y, KanMX, 2µ, bla, ColE1*	[116]

Table S5.2. Primers used for plasmid and strain construction in this study. *The following consensus is used for primer names: [fw/rv]_[Gene target]_[homology flank or part type]. Fw indicates a forward primer, Rv a reverse primer. Dg is a diagnostic primer.

# ID	Purpose*	Sequence
1693	Fw_pUD518	GCTGCTACTCATCCTAGTCC
3118	Rv_pUDC365	CTTGGCAGCAACAGGACTAGGATGAGTAGCAGCACGTTCC
10320	Fw_pUDC365	CATGCGCGGATGACACGAAC
10321	Rv_pUD518	CGTCGTGAGTTCGTGTCATC
16310	fw_DF_pt3aYTK	GCATCGTCTCATCGGTCTCATATGGCGCAATCGGTAGCGCTTCGAGCCCCCTACAGGGCTCCATTCTTGAGACCTGAGACGGCAT
16311	rv_DF_pt3aYTK	ATGCCGTCTCAGGTCTCAAGAATGGAGCCCTGTAGGGGGCTCGAAGCGCTACCGATTGCGCCATATGAGACCGATGAGACGATGC
16312	fw_D_pt3aYTK	GCATCGTCTCATCGGTCTCATATGGCGCAATCGGTAGCGCTTCTTGAGACCTGAGACGGCAT
16313	rv_D_pt3aYTK	ATGCCGTCTCAGGTCTCAAGAAGCGCTACCGATTGCGCCATATGAGACCGATGAGACGATGC
16314	fw_F_pt3aYTK	GCATCGTCTCATCGGTCTCATATGGAGCCCCCTACAGGGCTCTTTTCTTGAGACCTGAGACGGCAT
16315	rv_F_pt3aYTK	ATGCCGTCTCAGGTCTCAAGAAAAGAGCCCTGTAGGGGGCTCCATATGAGACCGATGAGACGATGC
16316	fw_h5SrRNA_pt3aYTK	GCATCGTCTCATCGGTCTCATATGGGCCTGGTTAGTACTTGGATGGGGGACCGCCAAGGAATACCGGGTGTTCTTGAGACCTGAGACGGCAT
16317	rv_h5SrRNA_pt3aYTK	ATGCCGTCTCAGGTCTCAAGAACACCCGGTATTCCTTGGCGGTCCCCCATCCAAGTACTAACCAGGCCCATATGAGACCGATGAGACGATGC
16318	fw_y5SrRNA_pt3aYTK	GCATCGTCTCATCGGTCTCATATGGACCGAGTAGTGTAGTGGGTGACCATACGCGAAACTCAGGTGTTCTTGAGACCTGAGACGGCAT
16319	rv_y5SrRNA_pt3aYTK	ATGCCGTCTCAGGTCTCAAGAACACCTGAGTTTCGCGTATGGTCACCCACTACACTACTCGGTCCATATGAGACCGATGAGACGATGC
16320	fw_hRP_pt3aYTK	GCATCGTCTCATCGGTCTCATATGTCTCCCTGAGCTTCAGGGAGTTCTTGAGACCTGAGACGGCAT
16321	rv_hRP_pt3aYTK	ATGCCGTCTCAGGTCTCAAGAACTCCCTGAAGCTCAGGGAGACATATGAGACCGATGAGACGATGC
16322	fw_yRP_pt3aYTK	GCATCGTCTCATCGGTCTCATATGACTGGGGAACCAGTTTCTTGAGACCTGAGACGGCAT
16323	rv_yRP_pt3aYTK	ATGCCGTCTCAGGTCTCAAGAAACTGGTTCCCCAGTCATATGAGACCGATGAGACGATGC
16324	fw_hMRP_pt3aYTK	GCATCGTCTCATCGGTCTCATATGAGAAGCGTATCCCGCTGAGCTTCTTGAGACCTGAGACGGCAT
16325	rv_hMRP_pt3aYTK	ATGCCGTCTCAGGTCTCAAGAAGCTCAGCGGGATACGCTTCTCATATGAGACCGATGAGACGATGC
16326	fw_yMRP_pt3aYTK	GCATCGTCTCATCGGTCTCATATGGTACCTCTATTGCAGGGTACTTCTTGAGACCTGAGACGGCAT
16327	rv_yMRP_pt3aYTK	ATGCCGTCTCAGGTCTCAAGAAGTACCCTGCAATAGAGGTACCATATGAGACCGATGAGACGATGC
16336	fw_D_pt4aYTK	GCATCGTCTCATCGGTCTCAATCCGCGCAATCGGTAGCGCTGGCTGAGACCTGAGACGGCAT
16337	rv_D_pt4aYTK	ATGCCGTCTCAGGTCTCAGCCAGCGCTACCGATTGCGCGGATTGAGACCGATGAGACGATGC
16338	fw_F_pt4aYTK	GCATCGTCTCATCGGTCTCAATCCGAGCCCCCTACAGGGCTCTTTGGCTGAGACCTGAGACGGCAT
16339	rv_F_pt4aYTK	ATGCCGTCTCAGGTCTCAGCCAAAGAGCCCTGTAGGGGGCTCGGATTGAGACCGATGAGACGATGC
16340	fw_h5SrRNA_pt4aYTK	GCATCGTCTCATCGGTCTCAATCCGGCCTGGTTAGTACTTGGATGGGGGACCGCCAAGGAATACCGGGTGTGGCTGAGACCTGAGACGGCAT
16341	rv_h5SrRNA_pt4aYTK	ATGCCGTCTCAGGTCTCAGCCACACCCGGTATTCCTTGGCGGTCCCCCATCCAAGTACTAACCAGGCCGGATTGAGACCGATGAGACGATGC
16342	fw_y5SrRNA_pt4aYTK	GCATCGTCTCATCGGTCTCAATCCGACCGAGTAGTGTAGTGGGTGACCATACGCGAAACTCAGGTGTGGCTGAGACCTGAGACGGCAT
16343	rv_y5SrRNA_pt4aYTK	ATGCCGTCTCAGGTCTCAGCCACACCTGAGTTTCGCGTATGGTCACCCACTACACTACTCGGTCGGATTGAGACCGATGAGACGATGC
16344	fw_hRP_pt4aYTK	GCATCGTCTCATCGGTCTCAATCCTCTCCCTGAGCTTCAGGGAGTGGCTGAGACCTGAGACGGCAT
16345	rv_hRP_pt4aYTK	ATGCCGTCTCAGGTCTCAGCCACTCCCTGAAGCTCAGGGAGAGGATTGAGACCGATGAGACGATGC

16346	fw_yRP_pt4aYTK	GCATCGTCTCATCGGTCTCAATCCACTGGGGAACCAGTTGGCTGAGACCTGAGACGGCAT
16347	rv_yRP_pt4aYTK	ATGCCGTCTCAGGTCTCAGCCAACTGGTTCCCCAGTGGATTGAGACCGATGAGACGATGC
16348	fw_hMRP_pt4aYTK	GCATCGTCTCATCGGTCTCAATCCAGAAGCGTATCCCGCTGAGCTGGCTGAGACCTGAGACGGCAT
16349	rv_hMRP_pt4aYTK	ATGCCGTCTCAGGTCTCAGCCAGCTCAGCGGGATACGCTTCTGGATTGAGACCGATGAGACGATGC
16350	fw_yMRP_pt4aYTK	GCATCGTCTCATCGGTCTCAATCCGTACCTCTATTGCAGGGTACTGGCTGAGACCTGAGACGGCAT
16351	rv_yMRP_pt4aYTK	ATGCCGTCTCAGGTCTCAGCCAGTACCCTGCAATAGAGGTACGGATTGAGACCGATGAGACGATGC
16352	fw_tracrRNA_pt4aYTK	GCATCGTCTCATCGGTCTCAATCCGTTTTAGAGCTAGAAATAGCAAG
16353	rv_tracrRNA_pt4aYTK	ATGCCGTCTCAGGTCTCAGCCAGCACCACCGACTCGGTG
16643	fw_mito-mTq2_YTKp3b	GCATCGTCTCATCGGTCTCATTCTATGGTATCAAAAGGTGAAG
16644	fw_mito-mTq2_YTKp3	GCATCGTCTCATCGGTCTCATATGGTATCAAAAGGTGAAG
16645	rv_mito-mTq2	ATGCCGTCTCAGGTCTCAGGATTTATTTATATAATTCATCCATACCTAATG
16837	Fw_pUDE929-HDV_YTK4	GATCGGTCTCAATCCGGCCGGCATGGTCCCAGCC
16838	Fw_pUDE929-HDV_YTK4b	GATCGGTCTCATGGCGGCCGGCATGGTCCCAGCC
16839	Rv_pUDE929-HH_YTK4	GATCGGTCTCACATAATCTAATGACGAGCTTACTCG
16840	Rv_mito-mTq2_pt3a	ATGCCGTCTCAGGTCTCAAGAATTATTTATATAATTCATCCATACCTAATG
16841	Fw_mito-mTq2-pt3a	GCATCGTCTCATCGGTCTCAATCCGTATCAAAAGGTGAAG
16842	Rv_mito-mTq2-pt4a	ATGCCGTCTCAGGTCTCAGCCATTATTTATATAATTCATCCATACCTAATG
18440	Fw_pUDC329_HDV	TGGGCAACACCTTCGGGTGGCGAATGGGACTAACTCGAGAGTGCTTTTAACTAAGAATTATTAGTCTTTTCTGC
18441	Rv_pUDC329_HH	GTTTCGTCCTCACGGACTCATCAGATGCGCATTTGTAATTAAAACTTAGATTAGATTGCTATGC
18442	Fw_mitoARG8-HH	GCGCATCTGATGAGTCCGTGAGG
18443	Rv_mitoARG8_HDV	GTCCCATTCGCCACCCGAAGGTGTTGC
18444	Fw_mitoARG8-dg	GGTGAAGGTGGTGTATTTCC
18445	Rv_mitoARG8-dg	AAAGCACCCATTGTTCTACC
18446	Fw_F1v2-pt3a	GGTCTCATATGGGTTCGAGCCCCCTACAGGGCTCCATTCTAGAGACC
18447	Rv_F1v2-pt3a	GGTCTCTAGAATGGAGCCCTGTAGGGGGCTCGAACCCATATGAGACC
18448	Fw_tRK2-tr93_pt3a	GGTCTCATATGGCCTTGTTAGCTCAGTTGGTAGAGCGTTCGGCTTTTAACCGAAATGTCAGGGGTTCGAGCCCCCTATGAGGCTCCATTCTAGAGACC
18449	Rv_tRK2-tr93_pt3a	GGTCTCTAGAATGGAGCCTCATAGGGGGCTCGAACCCCTGACATTTCGGTTAAAAGCCGAACGCTCTACCAACTGAGCTAACAAGGCCATATGAGACC
14022	Fw_pUDC286_YPRCtau3	AGAATGATTTACAATCTAGTCGCAAAAACAAGTACAGTGCTGACGTCCCATCTTTTAATGCATGCGCGGATGACACGAACTCA
14023	rv_pUDC286_YPRCtau3	CATTACCAATGAATGCTGTTTTTGCAGAAATAACGAGATATCTGCAATAAAAGCAAAAGTCCTTGATCTGTCGGGTGTCTGCT

References

1. Tröder, S.E. and Zevnik, B., History of genome editing: From meganucleases to CRISPR. *Laboratory Animals*, *2022*. ***56****(1): p. 60-68.*
2. Malina, C., Larsson, C., and Nielsen, J., Yeast mitochondria: an overview of mitochondrial biology and the potential of mitochondrial systems biology. *FEMS yeast research*, *2018*. ***18****(5).*
3. Gray, M.W., The endosymbiont hypothesis revisited, in International Review of Cytology, D.R. Wolstenholme and K.W. Jeon, Editors. 1992, *Academic Press*. p. 233-357.
4. Gray, M.W., Burger, G., and Lang, B.F., Mitochondrial evolution. *Science*, *1999*. ***283****(5407): p. 1476-1481.*
5. Foury, F., Roganti, T., Lecrenier, N., and Purnelle, B., The complete sequence of the mitochondrial genome of *Saccharomyces cerevisiae*. *FEBS letters*, *1998*. ***440****(3): p. 325-331.*
6. Schmidt, O., Pfanner, N., and Meisinger, C., Mitochondrial protein import: from proteomics to functional mechanisms. *Nature Reviews Molecular Cell Biology*, *2010*. ***11****(9): p. 655-667.*
7. Tuppen, H.A.L., Blakely, E.L., Turnbull, D.M., and Taylor, R.W., Mitochondrial DNA mutations and human disease. *Biochimica et Biophysica Acta (BBA) - Bioenergetics*, *2010*. ***1797****(2): p. 113-128.*
8. Silva-Pinheiro, P. and Minczuk, M., The potential of mitochondrial genome engineering. *Nature Reviews Genetics*, *2022*. ***23****(4): p. 199-214.*
9. Barrera-Paez, J.D. and Moraes, C.T., Mitochondrial genome engineering coming-of-age. *Trends in Genetics*, *2022*. ***38****(8): p. 869-880.*
10. Gammage, P.A., Moraes, C.T., and Minczuk, M., Mitochondrial genome engineering: the revolution may not be CRISPR-Ized. *Trends in Genetics*, *2018*. ***34****(2): p. 101-110.*
11. Yang, X., Jiang, J., Li, Z., Liang, J., and Xiang, Y., Strategies for mitochondrial gene editing. *Computational and Structural Biotechnology Journal*, *2021*. ***19****: p. 3319-3329.*
12. Pereira, C.V. and Moraes, C.T., Current strategies towards therapeutic manipulation of mtDNA heteroplasmy. *Frontiers in bioscience (Landmark edition)*, *2017*. ***22****(6): p. 991-1010.*
13. Mok, B.Y., de Moraes, M.H., Zeng, J., Bosch, D.E., Kotrys, A.V., Raguram, A., Hsu, F., Radey, M.C., Peterson, S.B., Mootha, V.K., Mougous, J.D., and Liu, D.R., A bacterial cytidine deaminase toxin enables CRISPR-free mitochondrial base editing. *Nature*, *2020*. ***583****(7817): p. 631-637.*
14. Bonnefoy, N. and Fox, T.D., Directed alteration of *Saccharomyces cerevisiae* mitochondrial DNA by biolistic transformation and homologous recombination, in Mitochondria: Practical Protocols, D. Leister and J.M. Herrmann, Editors. 2007, *Humana Press*: Totowa, NJ. p. 153-166.
15. Larosa, V. and Remacle, C., Transformation of the mitochondrial genome. *International Journal of Developmental Biology*, *2013*. ***57****(6-7-8): p. 659-665.*
16. Kim, H. and Kim, J.-S., A guide to genome engineering with programmable nucleases. *Nature Reviews Genetics*, *2014*. ***15****(5): p. 321-334.*
17. Chacinska, A., Koehler, C.M., Milenkovic, D., Lithgow, T., and Pfanner, N., Importing mitochondrial proteins: Machineries and mechanisms. *Cell*, *2009*. ***138****(4): p. 628-644.*
18. Schneider, A., Mitochondrial tRNA import and its consequences for mitochondrial translation. *Annual review of biochemistry*, *2011*. ***80****: p. 1033-1053.*
19. Li, K., Smagula, C., Parsons, W., Richardson, J., Gonzalez, M., Hagler, H., and Williams, R., Subcellular partitioning of MRP RNA assessed by ultrastructural and biochemical analysis. *Journal of Cell Biology*, *1994*. ***124****(6): p. 871-882.*
20. Doersen, C.J., Guerrier-Takada, C., Altman, S., and Attardi, G., Characterization of an RNase P activity from HeLa cell mitochondria. Comparison with the cytosol RNase P activity. *Journal of Biological Chemistry*, *1985*. ***260****(10): p. 5942-9.*
21. Magalhães, P.J., Andreu, A.L., and Schon, E.A., Evidence for the presence of 5S rRNA in mammalian mitochondria. *Molecular Biology of the Cell*, *1998*. ***9****(9): p. 2375-82.*
22. Tarassov, I., Entelis, N., and Martin, R.P., An intact protein translocating machinery is required for mitochondrial import of a yeast cytoplasmic tRNA. *Journal of molecular biology*, *1995*. ***245****(4): p. 315-23.*
23. Kolesnikova, O., Kazakova, H., Comte, C., Steinberg, S., Kamenski, P., Martin, R.P., Tarassov, I., and Entelis, N., Selection of RNA aptamers imported into yeast and human mitochondria. *RNA*, *2010*. ***16****(5): p. 926-41.*
24. Wang, G., Chen, H.-W., Oktay, Y., Zhang, J., Allen, E.L., Smith, G.M., Fan, K.C., Hong, J.S., French, S.W., and McCaffery, J.M.J.C., PNPASE regulates RNA import into mitochondria. *2010*. ***142****(3): p. 456-467.*
25. Wang, G., Shimada, E., Zhang, J., Hong, J.S., Smith, G.M., Teitell, M.A., and Koehler, C.M., Correcting human mitochondrial mutations with targeted RNA import. *Proceedings of the National Academy of Sciences*, *2012*. ***109****(13): p. 4840-4845.*
26. Zelenka, J., Alán, L., Jabůrek, M., and Ježek, P., Import of desired nucleic acid sequences using addressing motif of mitochondrial ribosomal 5S-rRNA for fluorescent *in vivo* hybridization of mitochondrial DNA and RNA. *Journal of bioenergetics and biomembranes*, *2014*. ***46****(2): p. 147-156.*
27. Kolesnikova, O., Entelis, N., Kazakova, H., Brandina, I., Martin, R.P., and Tarassov, I., Targeting of tRNA into yeast and human mitochondria: the role of anticodon nucleotides. *Mitochondrion*, *2002*. ***2****(1-2): p. 95-107.*
28. Loutre, R., Heckel, A.M., Smirnova, A., Entelis, N., and Tarassov, I., Can mitochondrial DNA be CRISPRized: Pro and contra. *IUBMB life*, *2018*. ***70****(12): p. 1233-1239.*
29. Hussain, S.A., Yalvac, M.E., Khoo, B., Eckardt, S., and McLaughlin, K.J., Adapting CRISPR/Cas9 system for targeting mitochondrial genome. *Frontiers in Genetics*, *2021*. ***12****: p. 627050.*
30. Bi, R., Li, Y., Xu, M., Zheng, Q., Zhang, D.-F., Li, X., Ma, G., Xiang, B., Zhu, X., Zhao, H., Huang, X., Zheng, P., and Yao, Y.-G., Direct evidence of CRISPR-Cas9-mediated mitochondrial genome editing. *The Innovation*, *2022*. ***3****(6): p. 100329.*
31. Bian, W.-P., Chen, Y.-L., Luo, J.-J., Wang, C., Xie, S.-L., and Pei, D.-S., Knock-in strategy for editing human and zebrafish mitochondrial DNA using mito-CRISPR/Cas9 system. *ACS synthetic biology*, *2019*. ***8****(4): p. 621-632.*
32. Amai, T., Tsuji, T., Ueda, M., and Kuroda, K., Development of a mito-CRISPR system for generating mitochondrial DNA-deleted strain in *Saccharomyces cerevisiae*. *Bioscience, Biotechnology, and Biochemistry*, *2020*. ***85****(4): p. 895-901.*
33. Antón, Z., Mullally, G., Ford, H.C., van der Kamp, M.W., Szczelkun, M.D., and Lane, J.D., Mitochondrial import, health and mtDNA copy number variability seen when using type II and type V CRISPR effectors. *Journal of Cell Science*, *2020*. ***133****(18).*
34. Schmiderer, L., Yudovich, D., Oburoglu, L., Hjort, M., and Larsson, J., Site-specific CRISPR-based mitochondrial DNA manipulation is limited by gRNA import. *Scientific reports*, *2022*. ***12****(1): p. 18687.*
35. Jeandard, D., Smirnova, A., Tarassov, I., Barrey, E., Smirnov, A., and Entelis, N., Import of non-coding RNAs into human mitochondria: A critical review and emerging approaches. *Cells*, *2019*. ***8****(3): p. 286.*
36. Sharon, E., Chen, S.-A.A., Khosla, N.M., Smith, J.D., Pritchard, J.K., and Fraser, H.B., Functional genetic variants revealed by massively parallel precise genome editing. *Cell*, *2018*. ***175****(2): p. 544-557. e16.*
37. Chandramouly, G., Zhao, J., McDevitt, S., Rusanov, T., Hoang, T., Borisonnik, N., Treddinick, T., Lopezcolorado, F.W., Kent, T., Siddique, L.A.,... Pomerantz, R.T., Polθ; reverse transcribes RNA and promotes RNA-templated DNA repair.

Science Advances, *2021*. ***7****(24): p. eabf1771*.

38. Storici, F., Bebenek, K., Kunkel, T.A., Gordenin, D.A., and Resnick, M.A., RNA-templated DNA repair. *Nature*, *2007*. ***447****(7142): p. 338-341*.
39. Baile, M.G. and Claypool, S.M., The power of yeast to model diseases of the powerhouse of the cell. *Frontiers in bioscience (Landmark edition)*, *2013*. ***18****: p. 241*.
40. Entian, K.-D. and Kötter, P., 25 Yeast Genetic Strain and Plasmid Collections, in Methods in Microbiology, I. Stansfield and M.J.R. Stark, Editors. 2007, *Academic Press*. p. 629-666.
41. Pronk, J.T., Auxotrophic yeast strains in fundamental and applied research. *Applied and environmental microbiology*, *2002*. ***68****(5): p. 2095-2100*.
42. Koster, C.C. and Daran-Lapujade, P., Arginine auxotrophy caused by deletion of the mitochondrial selectable marker *ARG8* in *Saccharomyces cerevisiae* can be reverted by loss-of-function mutations in *UME6*. *Chapter 4 of this thesis*.
43. Postma, E.D., Dashko, S., Lars, Shannara, K., Van Den Broek, M., Daran, J.-M., and Daran-Lapujade, P., A supernumerary designer chromosome for modular *in vivo* pathway assembly in *Saccharomyces cerevisiae*. *Nucleic acids research*, *2021*. ***49****(3): p. 1769-1783*.
44. Bouwknegt, J., Koster, C.C., Vos, A.M., Ortiz-Merino, R.A., Wassink, M., Luttik, M.A.H., van den Broek, M., Hagedoorn, P.L., and Pronk, J.T., Class-II dihydroorotate dehydrogenases from three phylogenetically distant fungi support anaerobic pyrimidine biosynthesis. *Fungal Biology and Biotechnology*, *2021*. ***8****(1): p. 10*.
45. Koster, C.C., Kleefeldt, A., van den Broek, M., Luttik, M., Daran, J.-M., and Daran-Lapujade, P., Long-read direct RNA sequencing of the mitochondrial transcriptome of *Saccharomyces cerevisiae* reveals condition-dependent intron turnover. *bioRxiv*, *2023: p. 2023.01. 19.524680*.
46. Gietz, R.D. and Woods, R.A., Transformation of yeast by lithium acetate/single-stranded carrier DNA/polyethylene glycol method, in Methods in enzymology. 2002, *Elsevier*. p. 87-96.
47. Lee, M.E., DeLoache, W.C., Cervantes, B., and Dueber, J.E., A highly characterized yeast toolkit for modular, multipart assembly. *ACS synthetic biology*, *2015*. ***4****(9): p. 975-986*.
48. Perli, T., Moonen, D.P.I., Broek, M.v.d., Pronk, J.T., and Daran, J.-M., Adaptive laboratory evolution and reverse engineering of single-vitamin prototrophies in *Saccharomyces cerevisiae*. *Applied and Environmental Microbiology*, *2020*. ***86****(12): p. e00388-20*.
49. Mans, R., Wijsman, M., Daran-Lapujade, P., and Daran, J.-M., A protocol for introduction of multiple genetic modifications in *Saccharomyces cerevisiae* using CRISPR/Cas9. *FEMS yeast research*, *2018*. ***18****(7): p. foy063*.
50. Schwabe, A. and Bruggeman, F.J., Single yeast cells vary in transcription activity not in delay time after a metabolic shift. *Nature communications*, *2014*. ***5****(1): p. 1-10*.
51. Schindelin, J., Arganda-Carreras, I., Frise, E., Kaynig, V., Longair, M., Pietzsch, T., Preibisch, S., Rueden, C., Saalfeld, S., Schmid, B.,... Cardona, A., Fiji: an open-source platform for biological-image analysis. *Nature methods*, *2012*. ***9****(7): p. 676-682*.
52. Boonekamp, F.J., Knibbe, E., Vieira-Lara, M.A., Wijsman, M., Luttik, M.A.H., van Eunen, K., den Ridder, M., Bron, R., Almonacid Suarez, A.M., van Rijn, P.,... Daran-Lapujade, P., Full humanization of the glycolytic pathway in *Saccharomyces cerevisiae*. *Cell reports*, *2022*. ***39****(13): p. 111010*.
53. de Hulster, E., Mooiman, C., Timmermans, R., and Mans, R., Automated evolutionary engineering of yeasts, in Yeast Metabolic Engineering. 2022, *Springer*. p. 255-270.
54. Bracher, J.M., de Hulster, E., Koster, C.C., van den Broek, M., Daran, J.-M.G., van Maris, A.J., and Pronk, J.T., Laboratory evolution of a biotin-requiring *Saccharomyces cerevisiae* strain for full biotin prototrophy and identification of causal mutations. *Applied and Environmental Microbiology*, *2017*. ***83****(16): p. e00892-17*.
55. Postma, E., Verduyn, C., Scheffers, W.A., and Van Dijken, J.P., Enzymic analysis of the crabtree effect in glucose-limited chemostat cultures of *Saccharomyces cerevisiae*. *Applied and environmental microbiology*, *1989*. ***55****(2): p. 468-477*.
56. Li, H. and Durbin, R., Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics*, *2009*. ***25****(14): p. 1754-60*.
57. Wronska, A.K., van den Broek, M., Perli, T., de Hulster, E., Pronk, J.T., and Daran, J.-M., Engineering oxygen-independent biotin biosynthesis in *Saccharomyces cerevisiae*. *Metabolic Engineering*, *2021*. ***67****: p. 88-103*.
58. Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., and Durbin, R., The Sequence Alignment/Map format and SAMtools. *Bioinformatics*, *2009*. ***25****(16): p. 2078-9*.
59. Walker, B.J., Abeel, T., Shea, T., Priest, M., Abouelliel, A., Sakthikumar, S., Cuomo, C.A., Zeng, Q., Wortman, J., Young, S.K., and Earl, A.M., Pilon: An integrated tool for comprehensive microbial variant detection and genome assembly improvement. *PloS one*, *2014*. ***9****(11): p. e112963*.
60. Robinson, J.T., Thorvaldsdóttir, H., Winckler, W., Guttman, M., Lander, E.S., Getz, G., and Mesirov, J.P., Integrative genomics viewer. *Nature biotechnology*, *2011*. ***29****(1): p. 24-6*.
61. Nijkamp, J.F., van den Broek, M.A., Geertman, J.-M.A., Reinders, M.J.T., Daran, J.-M.G., and de Ridder, D., *De novo* detection of copy number variation by co-assembly. Bioinformatics, 2012. **28**(24): p. 3195-3202.
62. den Ridder, M., van den Brandeler, W., Altiner, M., Daran-Lapujade, P., and Pabst, M., Proteome dynamics during transition from exponential to stationary phase under aerobic and anaerobic conditions in yeast. *bioRxiv*, *2022: p. 2022.09.23.509138*.
63. Ashburner, M., Ball, C.A., Blake, J.A., Botstein, D., Butler, H., Cherry, J.M., Davis, A.P., Dolinski, K., Dwight, S.S., Eppig, J.T.,... Sherlock, G., Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nature genetics*, *2000*. ***25****(1): p. 25-9*.
64. Boyle, E.I., Weng, S., Gollub, J., Jin, H., Botstein, D., Cherry, J.M., and Sherlock, G., GO::TermFinder—open source software for accessing Gene Ontology information and finding significantly enriched Gene Ontology terms associated with a list of genes. *Bioinformatics*, *2004*. ***20****(18): p. 3710-3715*.
65. Cherry, J.M., Hong, E.L., Amundsen, C., Balakrishnan, R., Binkley, G., Chan, E.T., Christie, K.R., Costanzo, M.C., Dwight, S.S., Engel, S.R.,... Wong, E.D., *Saccharomyces* Genome Database: the genomics resource of budding yeast. *Nucleic acids research*, *2012*. ***40****: p. D700-5*.
66. Weber-Lotfi, F. and Dietrich, A., Targeting therapeutic nucleic acids into mitochondria: A long challenge, in Mitochondrial biology and experimental therapeutics. 2018, *Springer*. p. 565-592.
67. Fox, T.D., Natural variation in the genetic code. *Annual review of genetics*, *1987*. ***21****(1): p. 67-91*.
68. Gagliardi, D., Stepien, P.P., Temperley, R.J., Lightowlers, R.N., and Chrzanowska-Lightowlers, Z.M.A., Messenger RNA stability in mitochondria: different means to an end. *Trends in Genetics*, *2004*. ***20****(6): p. 260-267*.
69. Gao, Y. and Zhao, Y., Self-processing of ribozyme-flanked RNAs into guide RNAs *in vitro* and *in vivo* for CRISPR-mediated genome editing. *Journal of Integrative Plant Biology*, *2014*. ***56****(4): p. 343-349*.
70. Salinas, T., Duchêne, A.-M., and Maréchal-Drouard, L., Recent advances in tRNA mitochondrial import. *Trends in biochemical sciences*, *2008*. ***33****(7): p. 320-329*.
71. Jauniaux, J.C., Urrestarazu, L.A., and Wiame, J.M., Arginine metabolism in *Saccharomyces cerevisiae*: subcellular localization of the enzymes. *Journal of bacteriology*, *1978*. ***133****(3): p. 1096-1107*.
72. Turk, E.M., Das, V., Seibert, R.D., and Andrulis, E.D., The mitochondrial RNA landscape of *Saccharomyces cerevisiae*. *PloS one*, *2013*. ***8****(10): p. e78105*.

73. Steele, D.F., Butler, C.A., and Fox, T.D., Expression of a recoded nuclear gene inserted into yeast mitochondrial DNA is limited by mRNA-specific translational activation. *Proceedings of the National Academy of Sciences*, *1996*. ***93****(11): p. 5253-5257*.
74. Mans, R., Daran, J.-M.G., and Pronk, J.T., Under pressure: evolutionary engineering of yeast strains for improved performance in fuels and chemicals production. *Current opinion in biotechnology*, *2018*. ***50****: p. 47-56*.
75. Visser, W., van Spronsen, E.A., Nanninga, N., Pronk, J.T., Kuenen, J.G., and van Dijken, J.P., Effects of growth conditions on mitochondrial morphology in *Saccharomyces cerevisiae*. *Antonie van Leeuwenhoek*, *1995*. ***67****(3): p. 243-253*.
76. Di Bartolomeo, F., Malina, C., Campbell, K., Mormino, M., Fuchs, J., Vorontsov, E., Gustafsson, C.M., and Nielsen, J., Absolute yeast mitochondrial proteome quantification reveals trade-off between biosynthesis and energy generation during diauxic shift. *Proceedings of the National Academy of Sciences*, *2020*. ***117****(13): p. 7524-7535*.
77. Oura, E., Effect of aeration on the growth energetics and biochemical composition of baker's yeast. 1972: Research Laboratories of the State Alcohol Monopoly.
78. Satroutdinov, A.D., Kuriyama, H., and Kobayashi, H., Oscillatory metabolism of *Saccharomyces cerevisiae* in continuous culture. *FEMS microbiology letters*, *1992*. ***98****(1-3): p. 261-267*.
79. Hans, M.A., Heinzle, E., and Wittmann, C., Free intracellular amino acid pools during autonomous oscillations in *Saccharomyces cerevisiae*. *Biotechnology and bioengineering*, *2003*. ***82****(2): p. 143-151*.
80. Chen, C.-I., McDonald, K.A., and Bisson, L., Oscillatory behavior of *Saccharomyces cerevisiae* in continuous culture: I. Effects of pH and nitrogen levels. *Biotechnology and bioengineering*, *1990*. ***36****(1): p. 19-27*.
81. Beuse, M., Kopmann, A., Diekmann, H., and Thoma, M., Oxygen, pH value, and carbon source induced changes of the mode of oscillation in synchronous continuous culture of *Saccharomyces cerevisiae*. *Biotechnology and bioengineering*, *1999*. ***63****(4): p. 410-417*.
82. Beuse, M., Bartling, R., Kopmann, A., Diekmann, H., and Thoma, M., Effect of the dilution rate on the mode of oscillation in continuous cultures of *Saccharomyces cerevisiae*. *Journal of Biotechnology*, *1998*. ***61****(1): p. 15-31*.
83. Brandberg, T., Gustafsson, L., and Franzén, C.J., The impact of severe nitrogen limitation and microaerobic conditions on extended continuous cultivations of *Saccharomyces cerevisiae* with cell recirculation. *Enzyme and microbial technology*, *2007*. ***40****(4): p. 585-593*.
84. Hochstrasser, M., Ubiquitin-dependent protein degradation. *Annual review of genetics*, *1996*. ***30****(1): p. 405-439*.
85. Dohmen, R.J., Willers, I., and Marques, A.J., Biting the hand that feeds: Rpn4-dependent feedback regulation of proteasome function. *Biochimica et Biophysica Acta (BBA) - Molecular Cell Research*, *2007*. ***1773****(11): p. 1599-1604*.
86. Nikko, E., Sullivan, J.A., and Pelham, H.R.B., Arrestin-like proteins mediate ubiquitination and endocytosis of the yeast metal transporter Smf1. *EMBO reports*, *2008*. ***9****(12): p. 1216-1221*.
87. Fang, N.N., Chan, G.T., Zhu, M., Comyn, S.A., Persaud, A., Deshaies, R.J., Rotin, D., Gsponer, J., and Mayor, T., Rsp5/Nedd4 is the main ubiquitin ligase that targets cytosolic misfolded proteins following heat stress. *Nature cell biology*, *2014*. ***16****(12): p. 1227-1237*.
88. Ivashov, V., Zimmer, J., Schwabl, S., Kahlhofer, J., Weys, S., Gstir, R., Jakschitz, T., Kremser, L., Bonn, G.K., Lindner, H., Huber, L.A., Leon, S., Schmidt, O., and Teis, D., Complementary α-arrestin-ubiquitin ligase complexes control nutrient transporter endocytosis in response to amino acids. *Elife*, *2020*. ***9***.
89. Wang, G., Yang, J., and Huibregtse, J.M., Functional Domains of the Rsp5 Ubiquitin-Protein Ligase. *Molecular and Cellular Biology*, *1999*. ***19****(1): p. 342-352*.
90. Kamenski, P., Smirnova, E., Kolesnikova, O., Krasheninnikov, I.A., Martin, R.P., Entelis, N., and Tarassov, I., tRNA mitochondrial import in yeast: Mapping of the import determinants in the carrier protein, the precursor of mitochondrial lysyl-tRNA synthetase. *Mitochondrion*, *2010*. ***10****(3): p. 284-293*.
91. Hussain, S.-R.A., Yalvac, M.E., Khoo, B., Eckardt, S., and McLaughlin, K.J., Adapting CRISPR/Cas9 system for targeting mitochondrial genome. *2020*.
92. Jo, A., Ham, S., Lee, G.H., Lee, Y.I., Kim, S., Lee, Y.S., Shin, J.H., and Lee, Y., Efficient mitochondrial genome editing by CRISPR/Cas9. *BioMed Research International*, *2015*. ***2015****: p. 305716*.
93. Furukawa, K., Innokentev, A., and Kanki, T., Regulatory mechanisms of mitochondrial autophagy: lessons from yeast. *Frontiers in Plant Science*, *2019*. ***10***.
94. O'Hara, R., Tedone, E., Ludlow, A., Huang, E., Arosio, B., Mari, D., and Shay, J.W., Quantitative mitochondrial DNA copy number determination using droplet digital PCR with single-cell resolution. *Genome research*, *2019*. ***29****(11): p. 1878-1888*.
95. Mercer, T.R., Neph, S., Dinger, M.E., Crawford, J., Smith, M.A., Shearwood, A.-M.J., Haugen, E., Bracken, C.P., Rackham, O., and Stamatoyannopoulos, J.A.J.C., The human mitochondrial transcriptome. *Cell*, *2011*. ***146****(4): p. 645-658*.
96. Gold, V.A., Chroscicki, P., Bragoszewski, P., and Chacinska, A., Visualization of cytosolic ribosomes on the surface of mitochondria by electron cryo-tomography. *EMBO reports*, *2017*. ***18****(10): p. 1786-1800*.
97. Michaud, M., Maréchal-Drouard, L., and Duchêne, A.-M., Targeting of cytosolic mRNA to mitochondria: Naked RNA can bind to the mitochondrial surface. *Biochimie*, *2014*. ***100****: p. 159-166*.
98. Saint-Georges, Y., Garcia, M., Delaveau, T., Jourdren, L., Le Crom, S., Lemoine, S., Tanty, V., Devaux, F., and Jacq, C., Yeast mitochondrial biogenesis: A role for the *PUF* RNA-binding protein Puf3p in mRNA localization. *PloS one*, *2008*. ***3****(6): p. e2293*.
99. Boldogh, I.R. and Pon, L.A., Purification and Subfractionation of Mitochondria from the Yeast *Saccharomyces cerevisiae*, in Methods in cell biology. 2007, *Academic Press*. p. 45-64.
100. Baker, M.J., Frazier, A.E., Gulbis, J.M., and Ryan, M.T., Mitochondrial protein-import machinery: correlating structure with function. *Trends in Cell biology*, *2007*. ***17****(9): p. 456-464*.
101. Tarassov, I., Entelis, N., and Martin, R., Mitochondrial import of a cytoplasmic lysine-tRNA in yeast is mediated by cooperation of cytoplasmic and mitochondrial lysyl-tRNA synthetases. *The EMBO journal*, *1995*. ***14****(14): p. 3461-3471*.
102. Smirnov, A., Comte, C., Mager-Heckel, A.M., Addis, V., Krasheninnikov, I.A., Martin, R.P., Entelis, N., and Tarassov, I., Mitochondrial enzyme rhodanese is essential for 5S ribosomal RNA import into human mitochondria. *Journal of Biological Chemistry*, *2010*. ***285****(40): p. 30792-803*.
103. Budin, I., Apel, A.R.R., Hummel, N.F., Mukhopadhyay, A., and Keasling, J.D., Methods for mitochondria and organelle genome editing. 2020, Google Patents.
104. Ford, H.C., Allen, W.J., Pereira, G.C., Liu, X., Dillingham, M.S., and Collinson, I., Towards a molecular mechanism underlying mitochondrial protein import through the TOM and TIM23 complexes. *Elife*, *2022*. ***11****: p. e75426*.
105. Gasteiger, E., Hoogland, C., Gattiker, A., Duvaud, S.e., Wilkins, M.R., Appel, R.D., and Bairoch, A., Protein identification and analysis tools on the ExPASy Server, in The Proteomics Protocols Handbook, J.M. Walker, Editor. 2005, *Humana Press*: Totowa, NJ. p. 571-607.
106. Lee, M., Choi, J.S., and Ko, K.S., Mitochondria targeting delivery of nucleic acids. *Expert Opinion on Drug Delivery*, *2008*. ***5****(8): p. 879-887*.
107. Coyne, H.J., III, Ciofi-Baffoni, S., Banci, L., Bertini, I., Zhang,

L., George, G.N., and Winge, D.R., The characterization and role of zinc binding in yeast Cox4 *Journal of Biological Chemistry*, *2007*. ***282****(12): p. 8926-8934*.

108. Veatch, J.R., McMurray, M.A., Nelson, Z.W., and Gottschling, D.E., Mitochondrial dysfunction leads to nuclear genome instability via an iron-sulfur cluster defect. *Cell*, *2009*. ***137****(7): p. 1247-58*.
109. Vowinckel, J., Hartl, J., Butler, R., and Ralser, M., MitoLoc: A method for the simultaneous quantification of mitochondrial network morphology and membrane potential in single cells. *Mitochondrion*, *2015*. ***24****: p. 77-86*.
110. Akiyoshi, B., Nelson, C.R., Duggan, N., Ceto, S., Ranish, J.A., and Biggins, S., The Mub1/Ubr2 ubiquitin ligase complex regulates the conserved Dsn1 kinetochore protein. *PLoS genetics*, *2013*. ***9****(2): p. e1003216*.
111. Herrero, E. and Thorpe, P.H., Synergistic control of kinetochore protein levels by Psh1 and Ubr2. *PLoS genetics*, *2016*. ***12****(2): p. e1005855*.
112. Marshall, R.S. and Vierstra, R.D., Dynamic regulation of the 26S proteasome: from synthesis to degradation. *Frontiers in Molecular Biosciences*, *2019*. ***6****: p. 40*.
113. Liu, C., Apodaca, J., Davis, L.E., and Rao, H., Proteasome inhibition in wild-type yeast *Saccharomyces cerevisiae* cells. *BioTechniques*, *2007*. ***42****(2): p. 158-162*.
114. Brandina, I., Smirnov, A., Kolesnikova, O., Entelis, N., Krasheninnikov, I.A., Martin, R.P., and Tarassov, I., tRNA import into yeast mitochondria is regulated by the ubiquitin-proteasome system. *FEBS letters*, *2007*. ***581****(22): p. 4248-4254*.
115. Mashego, M.R., Jansen, M.L.A., Vinke, J.L., van Gulik, W.M., and Heijnen, J.J., Changes in the metabolome of *Saccharomyces cerevisiae* associated with evolution in aerobic glucose-limited chemostats. *FEMS yeast research*, *2005*. ***5****(4-5): p. 419-430*.
116. Mans, R., van Rossum, H.M., Wijsman, M., Backx, A., Kuijpers, N.G., van den Broek, M., Daran-Lapujade, P., Pronk, J.T., van Maris, A.J., and Daran, J.-M.G., CRISPR/Cas9: a molecular Swiss army knife for simultaneous introduction of multiple genetic modifications in *Saccharomyces cerevisiae*. *FEMS yeast research*, *2015*. ***15****(2)*.
117. Goedhart, J., Von Stetten, D., Noirclerc-Savoye, M., Lelimousin, M., Joosen, L., Hink, M.A., Van Weeren, L., Gadella, T.W., and Royant, A., Structure-guided evolution of cyan fluorescent proteins towards a quantum yield of 93%. *Nature communications*, *2012*. ***3****(1): p. 1-9*.
118. Gruber, A.R., Lorenz, R., Bernhart, S.H., Neuböck, R., and Hofacker, I.L., The Vienna RNA Websuite. *Nucleic Acids Research*, *2008*. ***36****(suppl_2): p. W70-W74*.
119. Lorenz, R., Bernhart, S.H., Höner zu Siederdissen, C., Tafer, H., Flamm, C., Stadler, P.F., and Hofacker, I.L., ViennaRNA Package 2.0. *Algorithms for Molecular Biology*, *2011*. ***6****(1): p. 26*.
120. Esakova, O. and Krasilnikov, A.S., Of proteins and RNA: the RNase P/MRP family. *RNA*, *2010*. ***16****(9): p. 1725-1747*.

CHAPTER 6
CONCLUSIONS AND OUTLOOK

6.1 A critical evaluation of RNA import and processing in mitochondria

Despite the pressing need for methods to engineer the mitochondrial genome, from both a fundamental and therapeutic perspective, the development of the required tools is challenging. Although many targeted endonucleases (e.g., TALENs) are available, RNA-programmable endonucleases (RGENs) would be preferred, as they provide a fast and flexible method of genome engineering. RGENs require accessory RNA molecules to function, and the successful editing of the mitochondrial genome relies on the presence, of both the nuclease and an guiding RNA fragment in the mitochondrial matrix. In **Chapter 5,** it was shown that import to the mitochondrial matrix of RNA molecules with a size of 700 base pairs or longer could not be achieved in a reproducible manner. The efficiency of RNA import appeared similar irrespective of the import signal used, thus there was no apparent discrimination in the effectivity of RNA import signals. A similar effect has been observed for the import of shorter (100 nucleotide) gRNA [1-3], indicating that RNA import signals do not impact RNA import, regardless of the length of the cargo. This would imply that all cytosolic RNA could be potentially imported to the mitochondrial matrix. This is highly unlikely, as this would crowd the organelle with cytosolic RNA with no mitochondrial function.

Taken together, the data presented in **Chapter 5** calls for a better understanding of mitochondrial RNA import. Although parts of the import pathway of $tRNA^{Lys}$(CUU) are elucidated [4], the exact mechanism behind the role of pre-LysRS in RNA import is still unknown. It is still unclear if the protein and the RNA are associated upon import, or if pre-LysRS merely brings the RNA in proximity of the TOM complex and is then recognized by the TOM complex. Previous research indicates that the tRNA needs to undergo a conformational change to enter mitochondria, meaning that secondary structures may also impact the RNA import mechanisms [5]. Lastly, it is unclear how the negatively charged RNA molecule enters the heavily negatively charged mitochondrial matrix. In the case of DNA, it was demonstrated that the molecule can only enter the mitochondrial matrix when adhered to a strongly positively charged molecule [6]. This would indicate that RNA molecules that are able to enter the mitochondrial matrix are associated to a positively charged protein, such as pre-LysRS. The evolutionary engineering experiment presented in **Chapter 5** showed that RNA import cannot be simply achieved through a few mutations, and likely requires extensive changes in mitochondrial physiology, demonstrating the complexity of this phenotype.

More accurate screening methods for RNA import are required

Most studies describing mitochondrial RNA import relied solely on *in vitro* experiments. The isolation procedure of mitochondria alone can cause decoupling of mitochondria, resulting in rupture of the mitochondrial membranes and a subsequent loss of mitochondrial membrane potential. To further complicate the detection of RNA, multiple studies report the adhesion of cytosolic RNA to the mitochondrial surface [3, 7-11]. RNA import studies often rely on blotting techniques used on a mitochondrial subfraction of the cell. Isolated mitochondria, even in highly purified or mitoplast form, still contain contamination by cytosolic material; isolated mitochondrial protein fractions or transcriptomes were shown to still contain 20 – 80 % cytosolic proteins and RNA, respectively [8, 12]. Therefore, it could be argued that these sensitive techniques do not explicitly demonstrate the import of RNA, but rather RNase protection of the targeted RNA by association to the mitochondrial membrane [2]. In this light, the question arises if these studies specifically screen for mitochondrial uptake or for the association of RNA with mitochondrial proteins. The only demonstration of *in vivo* RNA import was achieved by Wang *et al.* [13] through the expression of human polynucleotide phosphorylase (PNPase). Whilst PNPase enabled *in vitro* and *in vivo* RNA import in multiple conditions, this finding remains highly disputed, as there is no consensus on the submitochondrial localization of the protein, either in the intermembrane space or the mitochondrial matrix, and because of its function as a key protein in the RNA degradasome [7, 14]. Concomitantly, the localization of 5S rRNA and predicted RNA components of mitochondrial RNases to the mitochondria in mammalian cells is disputed as well, and it is assumed to be an artefact of RNA species that associate with proteins or ribosomes on the mitochondrial membrane, or mRNA that encodes proteins that are co-translationally imported in the mitochondria [7].

To prevent any false positives, future studies assessing the RNA import *in vitro* or *in vivo* must carefully reconsider the screening method used to determine the level of RNA import and not rely on co-precipitation only. It has been widely and reproducibly demonstrated that recoding a gene with mitochondrial codons yields highly specific mitochondrial expression of a gene [15-24]. Hence, the strategy employed in **Chapter 5** could be further developed to demonstrate mitochondrial RNA import, given that any artifacts caused by fluorescence microscopy can be ruled out, and the strains cannot bypass the phenotype selected for, as was observed in **Chapter 4**. In the work performed in this research, it was assumed that the imported mRNA would be efficiently translated in the mitochondria. Although translation of the mito-mTq2 mRNA was verified by both fluorescence microscopy and Western blot, in both cases low levels were observed, and it is unclear if this localization is the result of inefficient RNA import and/or translation, or merely an artifact. If this screening approach is to be used in the future, more research should be done to characterize the expression of imported mRNA by the mitochondrial ribosomes, e.g., by *in vitro* transformation of isolated mitochondria with RNA [25]. Alternatively, a screening method not relying on mRNA expression, such as the use of fluorescent aptamers [26], would allow for screening of the import of shorter RNA without the need for their expression. The resolution and mitochondrial specificity of such a method could even be improved by e.g., Cas13 digestion of cytosolic aptamers [27]. Lastly, an import screening method could rely on harnessing the specificity of the mitochondrial RNA processing machinery. For example, an RNA molecule with internal specific mitochondrial RNA processing sites like a dodecamer, heptakaidecamer or I-SceI site [28-30] could be targeted to the mitochondria. Detection of the processed RNA fragments by sequencing or qPCR would confirm mitochondrial import without relying on translation of the molecule.

6.2 Towards full understanding of mitochondrial gene expression

Engineering even the most minimal forms of life still requires a thorough understanding of their gene expression and the regulation thereof. Although the 86 kb mitochondrial genome is relatively small, it has a complex mosaic structure and gene expression is controlled through various layers of RNA processing and splicing that are not yet fully understood. Disruption of splicing homeostasis can in some instances heavily alter mitochondrial physiology [31], suggesting that introns play an important role in regulating the level of expression of their respective genes. **Chapter 3** identifies a significant change in the level of spliced mitochondrial group II introns between respiro-fermentative growth on glucose and fully respiratory growth on ethanol. This may indicate a yet unforeseen role of these introns in regulation of mitochondrial genes in respiro-fermentative growth conditions. For example, the intron RNA sequences could regulate gene expression by interacting with the mtDNA or with other mtRNA species. Alternatively, the increased intron level could be the result of altered RNA turnover when cells are grown on glucose. Understanding the mechanisms of intron turnover and accumulation could shed new light on mechanisms of glucose repression and the Crabtree effect. Interestingly, there is a large heterogeneity in the presence of intron sequences in the mitochondrial genome within the *Saccharomyces* genus and even between *S. cerevisiae* strains [32, 33]. A large-scale genotype-phenotype screen on a variety on mitochondrial yeast genomes with different intron compositions and their respective mitochondrial intron processing and respiratory capacity could bring a wealth of new information on the regulation, essentiality and evolutionary history of the intron sequences and their role in mitochondrial physiology, energetics, and metabolism. At a clinical level, disruption of splicing homeostasis not only impacts cellular growth but can also lead to several mitochondrial pathologies and has implications in cellular aging in mammalian systems [34-36]. As *S. cerevisiae* is a popular model organism for mitochondrial functions, the sequencing method developed in **Chapter 3** could shed new light on the mitochondrial RNA-turnover in aging and disease. Additionally, readily developed methods could be used to characterize RNA turnover in the mammalian mitochondrial transcriptome [8].

6.3 The future of mitochondrial genome engineering

The data in **Chapter 5** characterizes mitochondrial import of RNA with a length 700 nucleotides and over. However, many of the studies attempting to express a CRISPR/Cas9 system in mitochondria rely on the import of a short gRNA linked to a targeting signal, which brings the length to an average of 100 to 150 nucleotides. With the simultaneous mitochondrial expression of Cas9 and a gRNA, these studies demonstrate depletion of (sections of) mtDNA [1, 37, 38]. A recent study by Schmiderer *et al.* [3], targeting a Cas9-base editor to the mitochondrial matrix, reported the occurrence of the expected editing event with a 1 % efficiency. However, their study uncovered that irrespective of the gRNA sequence used, the exact same base pair substitution at a single nucleotide position was observed. This indicates that the introduced mutation correlated to the presence of Cas9 in the mitochondria, but was independent the presence of a gRNA. In the light of these findings, the results of other controversial studies reporting Cas9-editing of the mitochondrial genome should be reevaluated, as these studies report mtDNA depletion rather than site specificity of the Cas9-gRNA complex, for example. Although mitochondrial CRISPR-studies generally report a 25 – 50 % reduction of mtDNA abundance in the presence of a gRNA [1, 37, 38], a DSB in the mtDNA can lead to general dysfunction of mitochondria, which may trigger maintenance systems reducing mtDNA such as mitophagy [39]. The observed reduction in mtDNA abundance might therefore be caused by other factors, including off-target activity of Cas9, an intrinsic effect of the already highly dynamic and variable mtDNA copy number [2, 40], or by prevention of replication of mtDNA by the docked Cas9 endonuclease [1]. Aside from mtDNA depletion, the integration of a DNA sequence would provide more solid evidence for gRNA import and Cas9 editing. Two studies report such a knock-in using Cas9 and an ssDNA repair: by PCR-amplification and sequencing these studies reported integration efficiencies of 0.03 – 0.23 % [41, 42]. On top of extremely low editing efficiencies and poorly reproducible data, neither of these studies were able to achieve a shift in mtDNA heteroplasmy, indicating that edits, if present at all, could not be maintained and that the developed genome editing method was not yet suitable for mtDNA engineering [43].

mtDNA engineering in mitochondrial disease

CRISPR/RGEN-based editing of the mitochondrial genome is mostly pursued for its applications in gene therapy, aiming at treating mitochondrial disorders that are rooted in mutations or deletions in the mtDNA. Over 250 pathogenic mtDNA genetic variants have been identified so far, and it is estimated that disease-causing mutations in the mtDNA have a prevalence of 1 in 5,000 humans, while effective treatments to cure these are not available [44, 45]. Several types of gene therapy treatments have been developed for human mitochondrial diseases, which do not necessarily depend on editing the human mitochondrial genome. Human mitochondrial defects are often heteroplasmic in nature, meaning that healthy and mutated copies of mtDNA co-exist within cells and the ratio of healthy and mutated alleles determines the degree of pathology [45, 46]. Many of these potential therapies focus on shifting or correcting the heteroplasmy towards the ‘healthy’ gene product, by e.g., attenuation of mutated proteins or aberrant mRNA through cytosolic expression and mitochondrial import of the corrected variants [43, 47-49]. More recently, promising mitochondrially targeted programmable nucleases such as mitoTALENs, mitoZFNs and mitochondrial base-editors have been developed, and have successfully been used to target, correct, or deplete heteroplasmic mitochondrial mutations [43, 49]. CRISPR-RGEN systems are much more flexible and precise, and unlike TALENs and ZFNs, do not require protein engineering to program the nuclease and can theoretically target any mutation [50, 51]. Therefore, establishing a functional mitoCRISPR-Cas9 is seen as the ‘holy grail’ of mitochondrial gene therapy. With the number of unanswered fundamental questions on nucleic acid import in the mitochondrial matrix and the observations that despite many efforts, there is no solid proof for Cas9 engineering of mitochondrial genomes yet, targeting of a gRNA to the mitochondrial matrix *in vivo*, without disrupting the mitochondria, may never be achieved in a clinical setting. Therefore, further development of the earlier described attenuation strategies for mitochondrial heteroplasmy or protein engineering for increased flexibility of the TALE- or ZFN-nucleases and base editors may be a more fruitful way to achieve gene therapy in mitochondrial disease. Alternatively, hereditary mitochondrial diseases can be targeted at the oocyte stage by transplantation of healthy mitochondrial genomes into mitochondria-depleted oocytes and subsequent *in vitro* fertilization [52, 53]. Mitochondrial replacement is now also considered

beyond germline engineering, as several studies reported a promising effectivity of therapies based on the transplantation of healthy mitochondria in tissues with mitochondrial defects. Thereby, mitochondrial transplantation could effectively rescue mitochondrial defects caused by ROS or mitochondrial mutations, with a minimal immune response and effective transfer and proliferation of healthy mitochondria [54, 55].

mtDNA engineering in synthetic biology

As mitochondrial RGEN editing is mostly developed in the scope of mitochondrial disease, in this context the depletion of mtDNA or base-substitution techniques suffice. However, in the scope of synthetic biology, the integration of genetic elements is required, which depends on the ability to import RNA or DNA. As described above, an increasing wealth of evidence suggests that RNA and DNA are not imported *in vivo* at a level that would sustain efficient engineering of homoplasmic mitochondrial genomes. Aside from building synthetic cells, the ability to engineer the mtDNA could have multiple applications in synthetic biology. For instance, this could be relevant for engineering efficient microbial cell factories, by engineering the compartmentalization of certain pathways while bypassing the varying efficiency of different MTS [56, 57], or the engineering of energy complexes for altered ATP conservation, cofactor supply or redox maintenance [58-61]. From a fundamental perspective, humanization of the yeast mitochondrial genome would offer a simplified and accessible model to study human mitochondrial diseases [62-64]. Mitochondrial genome engineering could also lead to fundamental insight in the endosymbiotic event that led to the existence of mitochondria by repeating or reversing the hypothetical DNA recombination events that occurred [65, 66], and into mitochondrial functions and gene expression. For example, it would allow for the removal of the group II introns that accumulate under growth on glucose, as described in **Chapter 3**, in order to study their physiological role. Alternatively, it would allow for efficient construction of (fluorescent) gene expression reporter systems, to study mitochondrial expression and regulation in different contexts, without the use of biolistic transformation [67]. Such systems may also lead to the discovery of novel mitochondrial elements that can be used as synthetic biology tools. For example, identification of the *ori* of the mtDNA of the yeast *Yarrowia lipolytica* lead to its application as an orthogonal origin of replication for maintenance of nuclear plasmids in this yeast [68].

In a synthetic biology context, the import of nucleic acids in the mitochondrial matrix may eventually be possible through additional modifications of mitochondria, something that is not possible when editing mtDNA in a therapeutic context. For example, virulence proteins that actively promote horizontal gene transfer by DNA-insertion in the nucleus could be modified to insert DNA into the mitochondria. This approach could utilize the proteins VirD2 and VirE2 of *Agrobacterium tumefaciens*, a plant pathogen that causes tumors in plants by transfer of oncogenic DNA to the plant nucleus. Whilst these proteins can be targeted to the mitochondrial matrix by addition of an MTS, successful DNA transfer to the mitochondria has not been demonstrated yet and likely requires further research [69, 70]. At the same time, other complexes involved in bacterial conjugation may also be explored and modified for expression on the mitochondrial membrane and selective DNA uptake, such as the T4S system [71, 72]. Nevertheless, the introduction of membrane pores is challenging, as strongly hydrophobic membrane proteins are not readily imported in mitochondria [73]. Additionally, the addition of pores may disturb maintenance of the mitochondrial membrane potential, impacting protein import altogether [74]. An alternative approach could be mimicking the RNA import mechanisms of *Trypanosoma* species, which import all tRNA into their mitochondria. Different protein complexes and mechanisms have been identified that may be involved in RNA import in these species, and it may be worthwhile to express the *Trypanosoma* mitochondrial membrane complexes in the yeast mitochondrial membrane to investigate if this enables RNA import [75-77]. Lastly, a strategy based on mitochondrial transplantation could also be employed for engineering and building of yeast mitochondrial genomes. Mitochondrial genomes can be transformed and engineered in hosts such as *E. coli* [78], or isolated mitochondria can be transformed *in vitro* with modified DNA through conjugation or electroporation [79, 80]. These modified isolated mitochondria could be inserted in a ρ^0 strain of *S. cerevisiae* by protoplast fusion [81]. Although such methods are more time-intensive that Cas9-mediated mtDNA engineering, they may be a more attainable approach in the short term.

6.4 Towards synthetic mitochondrial cells

Taken together, the ability to engineer mitochondrial genomes will have interesting applications in synthetic biology and it will be worthwhile to pursue the development of mitochondrial RGEN systems. As described, there are still several strategies for mitochondrial genome engineering using RGENs that could be investigated. However, in the scope of using mitochondria as synthetic cells, one might wonder if the number of technological challenges faced by mitochondrial genome engineering might not outweigh the advantages of using mitochondria as synthetic cell chassis.

Mitochondria: Not yet another top-down approach

The top-down construction of the first *Mycoplasma*-based synthetic cells has led to the construction of a minimal genome of ca. 531 kb, encoding 473 genes [82]. A surprizing insight gained from this research effort was that the function of over 30 % of these genes was, and is still undefined [83]. Although mitochondria, which contain approximately 1000 proteins, appear more complex, the organelle is in fact much better understood than even the most minimal synthetic cell to date. As a comparison, approximately 10 % of mitochondrial proteins have yet unknown function [84, 85]. As compared to bottom-up approaches, the most attractive aspect of mitochondria as chassis for synthetic cells is that spatial and self-organisation, maintenance, energy and redox conservation and division are not an issue, as the host yeast cell governs these. Initially, a full understanding of all mitochondrial functions is not required. Instead, single synthetic cell modules could be integrated on-by-one through stepwise replacement of the native mitochondrial functions and pathways, rather than 'booting up' an *in vitro* assembled cell all at once. By prolonged expressions of these modules in a mitochondrial context, they could even be evolved *in vivo*, to support the physiology of the synthetic mitochondria. This could be a complementary approach to *in vitro* evolution of synthetic cell components, which is likely required for a synthetic cell to fully integrate, but evolution of *in vitro* systems is more difficult to achieve [86]. Lastly, the susceptibility of yeast mitochondria to carbon catabolite repression may also be useful tool in synthetic cell construction, as it provides for modulation of mitochondrial morphology and DNA, RNA, and protein content. By providing the cell with different carbon sources, the level and exchange of DNA, and the level of transcription, translation and protein import in the mitochondrial synthetic cell can be regulated. For example, on glucose, the mitochondrial synthetic cells are elongated, and can fuse to exchange genetic information, while on ethanol, the mitochondrial synthetic cells will have a more cell-like phenotype, the DNA copy number can be increased and protein concentrations can be maximized.
In turn, expression and evolution of synthetic cell modules in mitochondria will increase the general understanding of mitochondria themselves. The unknown function of mitochondrial proteins can be elucidated through step-wise removal of mitochondrial proteins and complementing them by synthetic cell components and functions. For example, integrating orthogonal translation machinery in the mitochondria would allow for the systematic removal of (putative) components of the native mitochondrial translation machinery, thereby assessing their function without disrupting mitochondrial physiology. Therefore, using mitochondria as an intermediate for synthetic cells will offer research opportunities, not only in synthetic cell construction, but also in the general understanding of mitochondrial function.

The potential of assembling synthetic mitochondrial genomes

Despite mitochondria being promising starting points for synthetic cells, an efficient method for the integration of genes is lacking, and as described in **Chapter 5**, resolving this is a challenge on its own. Aside from mitochondrial genome engineering using RGENs, this study also explored methods towards building the genome for a synthetic cell *de novo* using the HR-machinery of *Saccharomyces cerevisiae*. As reviewed in **Chapter 2**, *in vivo* assembly can be used to generate DNA constructs up to several Mb and does not yet show a systematic limit in the size of the final construct or the number of fragments that can be assembled with this method. Therefore, the approaches discussed in this thesis could be integrated, where *in vivo* assembly using *S. cerevisiae* could be used to assemble fully synthetic and extended mitochondrial genomes. These genomes could be transformed into isolated mitochondria, and subsequently inserted in a ρ^0 strain of *S. cerevisiae* by protoplast fusion [81]. This would allow for complete genetic flexibility and extensive remodeling of the mitochondrial genome, bypassing the limitations in engineering the mtDNA

directly. Alternatively, any challenges with mitochondrial transformation could initially be overcome by nuclear expression of the synthetic mitochondrial genome and targeting the proteins to the mitochondrial matrix using an MTS. Such allotopic expression has only been achieved for *ATP8* and *COX2*, but is not straightforward, as the highly hydrophobic mitochondrial membrane proteins are not able to cross the inner membrane and are subject to degradation in the intermembrane space. Allotopic expression of these proteins therefore requires protein engineering to decrease their hydrophobicity while maintaining their function [73, 87, 88].

First steps towards synthetic mitochondrial cells

The proposed *in vivo* assembly of mitochondrial genomes may be the most fruitful way to engineer and build synthetic mtDNA with the knowledge we currently have. Employing this method, a first proof-of-principle experiment could be performed by re-integration of a nuclear-encoded mitochondrial gene in the mtDNA. For example, autonomous mitochondrial gene expression could be achieved by constitutive expression of the mitochondrial RNA polymerase Rpo41 and transcription factor Mtf1 on the mitochondrial genome [89], while removing these genes from the nuclear genome. Likewise, a simple orthogonal expression system, such as the T7 expression system, could be introduced in the mtDNA [90]. Another interesting approach would be to engineer mitochondria with an autonomous division cycle, as the core machinery for this process is relatively well-understood [91]. Alternatively, some of the synthetic cell modules developed within the BaSyC consortium could be expressed in a mitochondrial context, such as the minimal FtsZ-mediated or actomyosin-mediated membrane constriction [92, 93], deformation of the mitochondrial membranes [94], orthogonal genome replication using the bacteriophage Φ29 system [95], or alternative redox cofactor regeneration [96], which would constitute a first step towards a mitochondria-based synthetic cell. Lastly, assembly of synthetic mitochondrial genomes would allow for the extensive remodeling or *de novo* design of of mitochondrial genomes. This could include the assembly and transplantation of human mitochondrial genomes in yeast mitochondria, to model to study human mitochondrial disease [62-64].

Regardless of the method used to engineer synthetic cells or synthetic mitochondria, all of these exciting research opportunities could, as ambitioned by the BaSyC consortium, lead to the first bottom-up synthetic cell sometime in the current decade [97].

References

1. Loutre, R., Heckel, A.M., Smirnova, A., Entelis, N., and Tarassov, I., Can mitochondrial DNA be CRISPRized: Pro and contra. *IUBMB life*, *2018*. ***70****(12): p. 1233-1239.*
2. Antón, Z., Mullally, G., Ford, H.C., van der Kamp, M.W., Szczelkun, M.D., and Lane, J.D., Mitochondrial import, health and mtDNA copy number variability seen when using type II and type V CRISPR effectors. *Journal of Cell Science*, *2020*. ***133****(18).*
3. Schmiderer, L., Yudovich, D., Oburoglu, L., Hjort, M., and Larsson, J., Site-specific CRISPR-based mitochondrial DNA manipulation is limited by gRNA import. *Scientific reports*, *2022*. ***12****(1): p. 18687.*
4. Tarassov, I., Entelis, N., and Martin, R., Mitochondrial import of a cytoplasmic lysine-tRNA in yeast is mediated by cooperation of cytoplasmic and mitochondrial lysyl-tRNA synthetases. *The EMBO journal*, *1995*. ***14****(14): p. 3461-3471.*
5. Entelis, N.S., Kieffer, S., Kolesnikova, O.A., Martin, R.P., and Tarassov, I.A., Structural requirements of tRNALys for its import into yeast mitochondria. *Proceedings of the National Academy of Sciences*, *1998*. ***95****(6): p. 2838-2843.*
6. Niazi, A.K., Mileshina, D., Cosset, A., Val, R., Weber-Lotfi, F., and Dietrich, A., Targeting nucleic acids into mitochondria: progress and prospects. *Mitochondrion*, *2013*. ***13****(5): p. 548-558.*
7. Gammage, P.A., Moraes, C.T., and Minczuk, M., Mitochondrial genome engineering: the revolution may not be CRISPR-Ized. *Trends in Genetics*, *2018*. ***34****(2): p. 101-110.*
8. Mercer, T.R., Neph, S., Dinger, M.E., Crawford, J., Smith, M.A., Shearwood, A.-M.J., Haugen, E., Bracken, C.P., Rackham, O., and Stamatoyannopoulos, J.A.J.C., The human mitochondrial transcriptome. *Cell*, *2011*. ***146****(4): p. 645-658.*
9. Gold, V.A., Chroscicki, P., Bragoszewski, P., and Chacinska, A., Visualization of cytosolic ribosomes on the surface of mitochondria by electron cryo-tomography. *EMBO reports*, *2017*. ***18****(10): p. 1786-1800.*
10. Michaud, M., Maréchal-Drouard, L., and Duchêne, A.-M., Targeting of cytosolic mRNA to mitochondria: Naked RNA can bind to the mitochondrial surface. *Biochimie*, *2014*. ***100****: p. 159-166.*
11. Saint-Georges, Y., Garcia, M., Delaveau, T., Jourdren, L., Le Crom, S., Lemoine, S., Tanty, V., Devaux, F., and Jacq, C., Yeast mitochondrial biogenesis: A role for the *PUF* RNA-binding protein Puf3p in mRNA localization. *PloS one*, *2008*. ***3****(6): p. e2293.*
12. Boldogh, I.R. and Pon, L.A., Purification and subfractionation of mitochondria from the yeast *Saccharomyces cerevisiae*, in Methods in cell biology. 2007, *Academic Press*. p. 45-64.
13. Wang, G., Chen, H.-W., Oktay, Y., Zhang, J., Allen, E.L., Smith, G.M., Fan, K.C., Hong, J.S., French, S.W., and McCaffery, J.M., PNPase regulates RNA import into mitochondria. *Cell*, *2010*. ***142****(3): p. 456-467.*
14. Matilainen, S., Carroll, C.J., Richter, U., Euro, L., Pohjanpelto, M., Paetau, A., Isohanni, P., and Suomalainen, A., Defective mitochondrial RNA processing due to PNPT1 variants causes Leigh syndrome. *Human Molecular Genetics*, *2017*. ***26****(17): p. 3352-3361.*
15. Stenger, M., Le, D.T., Klecker, T., and Westermann, B., Systematic analysis of nuclear gene function in respiratory growth and expression of the mitochondrial genome in *S. cerevisiae*. *Microbial Cell*, *2020*. ***7****(9): p. 234-249.*
16. Steele, D.F., Butler, C.A., and Fox, T.D., Expression of a recoded nuclear gene inserted into yeast mitochondrial DNA is limited by mRNA-specific translational activation. *Proceedings of the National Academy of Sciences*, *1996*. ***93****(11): p. 5253-5257.*
17. Barros, M.H. and Tzagoloff, A., Aep3p-dependent translation of yeast mitochondrial *ATP8*. *Molecular Biology of the Cell*, *2017*. ***28****(11): p. 1426-1434.*
18. Bonnefoy, N. and Fox, T.D., *In vivo* analysis of mutated initiation codons in the mitochondrial *COX2* gene of *Saccharomyces cerevisiae* fused to the reporter gene *ARG8* reveals lack of downstream reinitiation. *Molecular and General Genetics MGG*, *2000*. ***262****(6): p. 1036-1046.*
19. Zhang, H., Chatterjee, A., and Singh, K.K., *Saccharomyces cerevisiae* Polymerase ζ Functions in Mitochondria. *Genetics*, *2006*. ***172****(4): p. 2683-2688.*
20. Rak, M., Tetaud, E., Godard, F., Sagot, I., Salin, B., Duvezin-Caubet, S., Slonimski, P.P., Rytka, J., and di Rago, J.-P., Yeast cells lacking the mitochondrial gene encoding the ATP synthase subunit 6 exhibit a selective loss of complex IV and unusual mitochondrial morphology *Journal of Biological Chemistry*, *2007*. ***282****(15): p. 10853-10864.*
21. Perez-Martinez, X., Broadley, S.A., and Fox, T.D., Mss51p promotes mitochondrial Cox1p synthesis and interacts with newly synthesized Cox1p. *The EMBO journal*, *2003*. ***22****(21): p. 5951-5961.*
22. De Silva, D., Poliquin, S., Zeng, R., Zamudio-Ochoa, A., Marrero, N., Perez-Martinez, X., Fontanesi, F., and Barrientos, A., The DEAD-box helicase Mss116 plays distinct roles in mitochondrial ribogenesis and mRNA-specific translation. *Nucleic acids research*, *2017*. ***45****(11): p. 6628-6643.*
23. Watts, T., Khalimonchuk, O., Wolf, R.Z., Turk, E.M., Mohr, G., and Winge, D.R., Mne1 is a novel component of the mitochondrial splicing apparatus responsible for processing of a *COX1* group I intron in yeast *Journal of Biological Chemistry*, *2011*. ***286****(12): p. 10137-10146.*
24. Williams, E.H. and Fox, T.D., Antagonistic signals within the *COX2* mRNA coding sequence control its translation in *Saccharomyces cerevisiae* mitochondria. *RNA*, *2003*. ***9****(4): p. 419-431.*
25. Wang, G., Shimada, E., Nili, M., Koehler, C.M., and Teitell, M.A., Mitochondria-targeted RNA import, in Mitochondrial Medicine. 2015, *Springer*. p. 107-116.
26. Nutiu, R. and Li, Y., Aptamers with fluorescence-signaling properties. *Methods*, *2005*. ***37****(1): p. 16-25.*
27. O'Connell, M.R., Molecular Mechanisms of RNA Targeting by Cas13-containing Type VI CRISPR–Cas Systems. *Journal of molecular biology*, *2019*. ***431****(1): p. 66-87.*
28. Osinga, K.A., De Vries, E., Van der Horst, G., and Tabak, H.F., Processing of yeast mitochondrial messenger RNAs at a conserved dodecamer sequence. *The EMBO journal*, *1984*. ***3****(4): p. 829-834.*
29. Turk, E.M., Das, V., Seibert, R.D., and Andrulis, E.D., The mitochondrial RNA landscape of *Saccharomyces cerevisiae*. *PloS one*, *2013*. ***8****(10): p. e78105.*
30. Colleaux, L., d'Auriol, L., Betermier, M., Cottarel, G., Jacquier, A., Galibert, F., and Dujon, B., Universal code equivalent of a yeast mitochondrial intron reading frame is expressed into *E. coli* as a specific double strand endonuclease. *Cell*, *1986*. ***44****(4): p. 521-33.*
31. Rudan, M., Dib, P.B., Musa, M., Kanunnikau, M., Sobočanec, S., Rueda, D., Warnecke, T., and Kriško, A., Normal mitochondrial function in *Saccharomyces cerevisiae* has become dependent on inefficient splicing. *Elife*, *2018*. ***7****: p. e35330.*
32. Wolters, J.F., Functional variation in the mitochondrial genome of the yeast *Saccharomyces cerevisiae*. 2018, Binghamton University - State University of New York.
33. Fonseca, P.L., De-Paula, R.B., Araújo, D.S., Tomé, L.M.R., Mendes-Pereira, T., Rodrigues, W.F.C., Del-Bem, L.-E., Aguiar, E.R., and Góes-Neto, A., Global characterization of fungal mitogenomes: new insights on genomic diversity and dynamism of coding genes and accessory elements. *Frontiers in Microbiology*, *2021: p. 3733.*
34. Heintz, C., Doktor, T.K., Lanjuin, A., Escoubas, C.C., Zhang, Y., Weir, H.J., Dutta, S., Silva-García, C.G., Bruun, G.H., Morantte, I., Hoxhaj, G., Manning, B.D., Andresen, B.S., and Mair, W.B., Splicing factor 1 modulates dietary restriction and TORC1 pathway longevity in C. elegans. *Nature*, *2017*.

541*(7635): p. 102-106.*

35. Wang, G.-S. and Cooper, T.A., Splicing in disease: disruption of the splicing code and the decoding machinery. *Nature Reviews Genetics*, *2007.* ***8****(10): p. 749-761.*
36. Chan, D.C., Mitochondrial dynamics and its involvement in disease. *Annual review of pathology: mechanisms of disease*, *2020.* ***15****: p. 235-259.*
37. Hussain, S.A., Yalvac, M.E., Khoo, B., Eckardt, S., and McLaughlin, K.J., Adapting CRISPR/Cas9 system for targeting mitochondrial genome. *Frontiers in Genetics*, *2021.* ***12****: p. 627050.*
38. Jo, A., Ham, S., Lee, G.H., Lee, Y.I., Kim, S., Lee, Y.S., Shin, J.H., and Lee, Y., Efficient mitochondrial genome editing by CRISPR/Cas9. *BioMed Research International*, *2015.* ***2015****: p. 305716.*
39. Furukawa, K., Innokentev, A., and Kanki, T., Regulatory mechanisms of mitochondrial autophagy: lessons from yeast. *Frontiers in Plant Science*, *2019.* ***10****.*
40. O'Hara, R., Tedone, E., Ludlow, A., Huang, E., Arosio, B., Mari, D., and Shay, J.W., Quantitative mitochondrial DNA copy number determination using droplet digital PCR with single-cell resolution. *Genome research*, *2019.* ***29****(11): p. 1878-1888.*
41. Bi, R., Li, Y., Xu, M., Zheng, Q., Zhang, D.-F., Li, X., Ma, G., Xiang, B., Zhu, X., Zhao, H., Huang, X., Zheng, P., and Yao, Y.-G., Direct evidence of CRISPR-Cas9-mediated mitochondrial genome editing. *The Innovation*, *2022.* ***3****(6): p. 100329.*
42. Bian, W.-P., Chen, Y.-L., Luo, J.-J., Wang, C., Xie, S.-L., and Pei, D.-S., Knock-in strategy for editing human and zebrafish mitochondrial DNA using mito-CRISPR/Cas9 system. *ACS synthetic biology*, *2019.* ***8****(4): p. 621-632.*
43. Barrera-Paez, J.D. and Moraes, C.T., Mitochondrial genome engineering coming-of-age. *Trends in Genetics*, *2022.* ***38****(8): p. 869-880.*
44. Gorman, G.S., Schaefer, A.M., Ng, Y., Gomez, N., Blakely, E.L., Alston, C.L., Feeney, C., Horvath, R., Yu-Wai-Man, P., Chinnery, P.F., Taylor, R.W., Turnbull, D.M., and McFarland, R., Prevalence of nuclear and mitochondrial DNA mutations related to adult mitochondrial disease. *Annals of Neurology*, *2015.* ***77****(5): p. 753-759.*
45. Tuppen, H.A.L., Blakely, E.L., Turnbull, D.M., and Taylor, R.W., Mitochondrial DNA mutations and human disease. *Biochimica et Biophysica Acta (BBA) - Bioenergetics*, *2010.* ***1797****(2): p. 113-128.*
46. Zhu, Z. and Wang, X., Significance of mitochondria DNA mutations in diseases, in Mitochondrial DNA and Diseases, H. Sun and X. Wang, Editors. 2017, *Springer Singapore*: Singapore. p. 219-230.
47. Kolesnikova, O.A., Entelis, N.S., Jacquin-Becker, C., Goltzene, F., Chrzanowska-Lightowlers, Z.M., Lightowlers, R.N., Martin, R.P., and Tarassov, I., Nuclear DNA-encoded tRNAs targeted into mitochondria can rescue a mitochondrial DNA mutation associated with the MERRF syndrome in cultured human cells. *Human molecular genetics*, *2004.* ***13****(20): p. 2519-2534.*
48. Chernega, T., Choi, J., Salmena, L., and Andreazza, A.C., Mitochondrion-targeted RNA therapies as a potential treatment strategy for mitochondrial diseases. *Molecular Therapy - Nucleic Acids*, *2022.* ***30****: p. 359-377.*
49. Soldatov, V.O., Kubekina, M.V., Skorkina, M.Y., Belykh, A.E., Egorova, T.V., Korokin, M.V., Pokrovskiy, M.V., Deykin, A.V., and Angelova, P.R., Current advances in gene therapy of mitochondrial diseases. *Journal of Translational Medicine*, *2022.* ***20****(1): p. 562.*
50. Silva-Pinheiro, P. and Minczuk, M., The potential of mitochondrial genome engineering. *Nature Reviews Genetics*, *2022.* ***23****(4): p. 199-214.*
51. Yang, X., Jiang, J., Li, Z., Liang, J., and Xiang, Y., Strategies for mitochondrial gene editing. *Computational and Structural Biotechnology Journal*, *2021.* ***19****: p. 3319-3329.*
52. Tachibana, M., Sparman, M., Sritanaudomchai, H., Ma, H., Clepper, L., Woodward, J., Li, Y., Ramsey, C., Kolotushkina, O., and Mitalipov, S., Mitochondrial gene replacement in primate offspring and embryonic stem cells. *Nature*, *2009.* ***461****(7262): p. 367-372.*
53. Sharma, H., Singh, D., Mahant, A., Sohal, S.K., Kesavan, A.K., and Samiksha, Development of mitochondrial replacement therapy: A review. *Heliyon*, *2020.* ***6****(9): p. e04643.*
54. Gollihue, J.L. and Rabchevsky, A.G., Prospects for therapeutic mitochondrial transplantation. *Mitochondrion*, *2017.* ***35****: p. 70-79.*
55. Park, A., Oh, M., Lee, S.J., Oh, K.-J., Lee, E.-W., Lee, S.C., Bae, K.-H., Han, B.S., and Kim, W.K., Mitochondrial transplantation as a novel therapeutic strategy for mitochondrial diseases. *International Journal of Molecular Sciences*, *2021.* ***22****(9): p. 4793.*
56. Dong, C., Shi, Z., Huang, L., Zhao, H., Xu, Z., and Lian, J., Cloning and characterization of a panel of mitochondrial targeting sequences for compartmentalization engineering in *Saccharomyces cerevisiae*. *Biotechnology and bioengineering*, *2021.*
57. Hammer, S.K., Zhang, Y., and Avalos, J.L., Mitochondrial compartmentalization confers specificity to the 2-ketoacid recursive pathway: Increasing isopentanol production in *Saccharomyces cerevisiae*. *ACS synthetic biology*, *2020.* ***9****(3): p. 546-555.*
58. Juergens, H., Hakkaart, X.D.V., Bras, J.E., Vente, A., Wu, L., Benjamin, K.R., Pronk, J.T., Daran-Lapujade, P., and Mans, R., Contribution of complex I NADH dehydrogenase to respiratory energy coupling in glucose-grown cultures of *Ogataea parapolymorpha*. *Applied and environmental microbiology*, *2020.* ***86****(15): p. e00678-20.*
59. de Kok, S., Kozak, B.U., Pronk, J.T., and van Maris, A.J., Energy coupling in *Saccharomyces cerevisiae*: selected opportunities for metabolic engineering. *FEMS yeast research*, *2012.* ***12****(4): p. 387-97.*
60. Chen, R., Gao, J., Yu, W., Chen, X., Zhai, X., Chen, Y., Zhang, L., and Zhou, Y.J., Engineering cofactor supply and recycling to drive phenolic acid biosynthesis in yeast. *Nature Chemical Biology*, *2022.* ***18****(5): p. 520-529.*
61. Cardenas, J. and Da Silva, N.A., Engineering cofactor and transport mechanisms in *Saccharomyces cerevisiae* for enhanced acetyl-CoA and polyketide biosynthesis. *Metabolic Engineering*, *2016.* ***36****: p. 80-89.*
62. Roohvand, F., Ehsani, P., Abdollahpour-Alitappeh, M., Shokri, M., and Kossari, N., Biomedical applications of yeasts - a patent view, part two: era of humanized yeasts and expanded applications. *Expert Opinion on Therapeutic Patents*, *2020.* ***30****(8): p. 609-631.*
63. Kachroo, A.H., Vandeloo, M., Greco, B.M., and Abdullah, M., Humanized yeast to model human biology, disease and evolution. *Disease Models & Mechanisms*, *2022.* ***15****(6).*
64. Laurent, J.M., Young, J.H., Kachroo, A.H., and Marcotte, E.M., Efforts to make and apply humanized yeast. *Briefings in Functional Genomics*, *2016.* ***15****(2): p. 155-163.*
65. Kleine, T., Maier, U.G., and Leister, D., DNA transfer from organelles to the nucleus: the idiosyncratic genetics of endosymbiosis. *Annual Review of Plant Biology*, *2009.* ***60****: p. 115-38.*
66. Timmis, J.N., Ayliffe, M.A., Huang, C.Y., and Martin, W., Endosymbiotic gene transfer: organelle genomes forge eukaryotic chromosomes. *Nature Reviews Genetics*, *2004.* ***5****(2): p. 123-135.*
67. Suhm, T., Habernig, L., Rzepka, M., Kaimal, J.M., Andreasson, C., Buttner, S., and Ott, M., A novel system to monitor mitochondrial translation in yeast. *Microbial Cell*, *2018.* ***5****(3): p. 158-164.*
68. Cui, Z., Zheng, H., Jiang, Z., Wang, Z., Hou, J., Wang, Q., Liang, Q., and Qi, Q., Identification and characterization of the mitochondrial replication origin for stable and episomal expression in *Yarrowia lipolytica*. *ACS synthetic biology*, *2021.* ***10****(4): p. 826-835.*
69. Budin, I., Apel, A.R.R., Hummel, N.F., Mukhopadhyay, A.,

and Keasling, J.D., Methods for mitochondria and organelle genome editing. 2020, Google Patents.

70. Roushan, M.R., Visualization of effector protein translocation from *Agrobacterium tumefaciens* into host cells. 2018, Leiden University.
71. Llosa, M. and de la Cruz, F., Bacterial conjugation: a potential tool for genomic engineering. *Research in Microbiology*, *2005*. ***156****(1): p. 1-6*.
72. Koraimann, G., Spread and persistence of virulence and antibiotic resistance genes: A ride on the F plasmid conjugation module. *EcoSal Plus*, *2018*. ***8****(1)*.
73. Artika, I.M., Allotopic expression of mitochondrial genes: Basic strategy and progress. *Genes & Diseases*, *2020*. ***7****(4): p. 578-584*.
74. Moulin, C., Caumont-Sarcos, A., and Ieva, R., Mitochondrial presequence import: Multiple regulatory knobs fine-tune mitochondrial biogenesis and homeostasis. *Biochimica et Biophysica Acta (BBA) - Molecular Cell Research*, *2019*. ***1866****(5): p. 930-944*.
75. Seidman, D., Johnson, D., Gerbasi, V., Golden, D., Orlando, R., and Hajduk, S., Mitochondrial membrane complex that contains proteins necessary for tRNA import in *Trypanosoma brucei*. *Journal of Biological Chemistry*, *2012*. ***287****(12): p. 8892-8903*.
76. Tschopp, F., Charrière, F., and Schneider, A., *In vivo* study in *Trypanosoma brucei* links mitochondrial transfer RNA import to mitochondrial protein import. *EMBO reports*, *2011*. ***12****(8): p. 825-32*.
77. Hancock, K. and Hajduk, S.L., The mitochondrial tRNAs of *Trypanosoma brucei* are nuclear encoded. *Journal of Biological Chemistry*, *1990*. ***265****(31): p. 19208-15*.
78. Yoon, Y.G. and Koob, M.D., Efficient cloning and engineering of entire mitochondrial genomes in *Escherichia coli* and transfer into transcriptionally active mitochondria. *Nucleic acids research*, *2003*. ***31****(5): p. 1407-15*.
79. Collombet, J.-M., Wheeler, V.C., Vogel, F., and Coutelle, C., Introduction of plasmid DNA into isolated mitochondria by electroporation: A novel approach toward gene correction for mitochondrial disorders *Journal of Biological Chemistry*, *1997*. ***272****(8): p. 5342-5347*.
80. Yoon, Y.G. and Koob, M.D., Transformation of isolated mammalian mitochondria by bacterial conjugation. *Nucleic Acids Research*, *2005*. ***33****(16): p. e139-e139*.
81. Fukuda, H. and Kimura, A., Transfer of mitochondria into protoplasts of *Saccharomyces cerevisiae* by mini-protoplast fusion. *FEBS letters*, *1980*. ***113****(1): p. 58-60*.
82. Glass, J.I., Assad-Garcia, N., Alperovich, N., Yooseph, S., Lewis, M.R., Maruf, M., Hutchison, C.A., Smith, H.O., and Venter, J.C., Essential genes of a minimal bacterium. *Proceedings of the National Academy of Sciences*, *2006*. ***103****(2): p. 425-430*.
83. Hutchison, C.A., Chuang, R.-Y., Noskov, V.N., Assad-Garcia, N., Deerinck, T.J., Ellisman, M.H., Gill, J., Kannan, K., Karas, B.J., and Ma, L., Design and synthesis of a minimal bacterial genome. *Science*, *2016*. ***351****(6280): p. aad6253*.
84. Di Bartolomeo, F., Malina, C., Campbell, K., Mormino, M., Fuchs, J., Vorontsov, E., Gustafsson, C.M., and Nielsen, J., Absolute yeast mitochondrial proteome quantification reveals trade-off between biosynthesis and energy generation during diauxic shift. *Proceedings of the National Academy of Sciences*, *2020*. ***117****(13): p. 7524-7535*.
85. Gonczarowska-Jorge, H., Zahedi, R.P., and Sickmann, A., The proteome of baker's yeast mitochondria. *Mitochondrion*, *2017*. ***33****: p. 15-21*.
86. Abil, Z. and Danelon, C., Roadmap to building a cell: An evolutionary approach. *Frontiers in Bioengineering and Biotechnology*, *2020*. ***8****: p. 927*.
87. Gearing, D.P., McMullen, G.L., and Nagley, P., Chemical synthesis of a mitochondrial gene designed for expression in the yeast nucleus. *Biochemistry international*, *1985*. ***10****(6): p. 907-915*.
88. Rubalcava-Gracia, D., García-Rincón, J., Pérez-Montfort, R., Hamel, P.P., and González-Halphen, D., Key within-membrane residues and precursor dosage impact the allotopic expression of yeast subunit II of cytochrome *c* oxidase. *Molecular Biology of the Cell*, *2019*. ***30****(18): p. 2358-2366*.
89. Contamine, V. and Picard, M., Maintenance and integrity of the mitochondrial genome: A plethora of nuclear genes in the budding yeast. *Microbiology and Molecular Biology Reviews*, *2000*. ***64****(2): p. 281-315*.
90. Pinkham, J.L., Dudley, A.M., and Mason, T.L., T7 RNA polymerase-dependent expression of *COXII* in yeast mitochondria. *Molecular and Cellular Biology*, *1994*. ***14****(7): p. 4643-52*.
91. Imoto, Y., Itoh, K., and Fujiki, Y., Molecular basis of mitochondrial and peroxisomal division machineries. *International Journal of Molecular Sciences*, *2020*. ***21****(15)*.
92. Godino, E., López, J.N., Zarguit, I., Doerr, A., Jimenez, M., Rivas, G., and Danelon, C., Cell-free biogenesis of bacterial division proto-rings that can constrict liposomes. *Communications Biology*, *2020*. ***3****(1): p. 539*.
93. Baldauf, L., van Buren, L., Fanalista, F., and Koenderink, G.H., Actomyosin-driven division of a synthetic cell. *ACS synthetic biology*, *2022*. ***11****(10): p. 3120-3133*.
94. De Franceschi, N., Pezeshkian, W., Fragasso, A., Bruininks, B.M.H., Tsai, S., Marrink, S.J., and Dekker, C., Synthetic membrane shaper for controlled liposome deformation. *ACS Nano*, *2023*. ***17****(2): p. 966-978*.
95. van Nies, P., Westerlaken, I., Blanken, D., Salas, M., Mencia, M., and Danelon, C., Self-replication of DNA by its encoded proteins in liposome-based synthetic cells. *Nat Commun*, *2018*. ***9****(1): p. 1583*.
96. Partipilo, M., Ewins, E.J., Fralliccardi, J., Robinson, T., Poolman, B., and Slotboom, D.J., Minimal pathway for the regeneration of redox cofactors. *JACS Au*, *2021*. ***1****(12): p. 2280-2293*.
97. Powell, K., Biology from scratch. *Nature*, *2018*. ***563****: p. 172-175*.

ACKNOWLEDGEMENTS

I have always joked that the acknowledgements section would be the thickest chapter of my thesis – partly because I was afraid of never obtaining results, but mostly because of the amazing people who have been there for me the last 5 years, and I could not imagine how many words I would need to properly thank them for what they have meant to me. Although this section is in the end not the thickest chapter, it is the most important (and likely most well-read) chapter, as it summarizes the collaborations, friendships and relationships that have developed over my PhD, and it is because of them and their support that this book is now in your hands or on your screen. Regardless, it wouldn't be anything like me to keep it short; since all of you invested so much time and effort in me, I want to at least thank all of you properly!

The biggest thank you obviously goes to my promotor and daily supervisor **Pascale**. My PhD experience hasn't always been an easy ride, both scientifically and personally, but for the past years I have always trusted and valued your support. The project has been a challenge from the start, setting up new research lines and venturing into new, challenging and sometimes seemingly impossible fields of science is not for the faint-hearted and it was a roller-coaster of emotions, especially figuring out if the RNA was or was not in the mitochondria: I guess we will never know for sure. Despite the constant frustration of not knowing how to force RNA into mitochondria, it mostly it has been a fun and exciting experience and somehow even deepened my enthusiasm for science. Thank you for fully trusting me with the project: I have always felt the freedom to do cool science, but trusting you were there to back me up. You taught me a lot, not only how to do science, being critical of results (including all controls ☺), but also how to navigate and withstand the (sometimes harsh) academic environment as a (female) scientist. I have really enjoyed our collaboration, meetings, as well as sharing our latest artistic creations, and am happy to have had you as a mentor. I'm curious whether the new mitochondrial research lines lead somewhere, and we will surely stay in touch!

Another big thank you to my other promotor **Jack**, even though you were maybe less involved in my thesis than planned, you were there at the critical moments when I needed you. Despite your busy schedule running the department, you were always able to make time for important matters, and I really appreciate the time you made for good advice, both personal, career-wise and science-wise. I have always felt backed-up and valued by you, I really appreciate your support for my science communication work, and providing me with opportunities to do so. I am very happy to have collaborated with a scientist like you, who is friendly, passionate and always working hard to make the department a welcoming place to do cool science!

Jean-Marc, even though you are not one of my 'official' supervisors, off the record you have been a great driving force behind the projects and you have been of great value to my work. Your seemingly never-ending wealth of knowledge on sequencing, amino acids, yeast physiology and molecular biology (including who published what and when) was a tremendous help in brainstorming about setting up and designing experiments, data analysis and finding creative solutions to experimental problems, sometimes even leading to full standalone chapters ("*Why not just Nanopore the mitochondria if you want to detect the RNA*?")? Secondly, as our group leader, I think you do a great job in keeping IMB a happy, well-functioning and fun working environment!

Marijke, I am confident that without you by my side, this thesis would not have been here. Firstly because you have helped me with almost every experiment presented in this thesis. You should have been part of my materials and methods section, but sadly that is frowned upon and not deemed reproducible in scientific publications. All jokes aside, you have been a great friend, mentor and support for me the past years. You taught me all the important skills and what you and I didn't know, we figured out together: sensitive microscopy, isolating boat-loads of mitochondria, new enzyme assays, as long as the day started with a cup of coffee to go through the plan, and ended with chocolate peanuts, it was going to be all right. Thank you for all the laughs we shared, the mistakes we shared (almost throwing away isolated mitochondria, really throwing away the isolated nuclei), the serious talks while dissolving sticky pellets, the fun art projects and generally always being there for me, for experiments, talks or hugs. You are an amazing and strong woman and am happy to have a friend like you, and to have you by my side once more when I defend my thesis, I wouldn't want it otherwise!

Nicolò, you are one of the kindest and most fun people I know. Having you as a friend alongside me in the lab has definitely made the PhD experience unforgettable. Thank you for all the fun moments: the dinners at Hummus, shopping sprees at TK Maxx, hours of giggling about labelling the single stranded DNA tubes, your endless supply of memes, pulling pranks on students or coworkers, but also the serious times: talks about being weak, talks about being strong, I have definitely learned from you how to face the PhD struggle but to also how to laugh your way through it. I still miss having you around in the lab but I am happy we live in the same city again! I was honored to be your paranymph and I am again honored you accepted to be mine, although I am terrified for the cabaret you will put together…

The cornerstones of any thesis, including mine, are of course our wonderful technicians and staff; without them there would be no IMB, but simultaneously they all happen to be wonderful people who made my time at IMB great. **Erik**, I know I entered the lab as a loudmouth who would 'never ever' run a bioreactor, look where that got me… Thank you for teaching me all the ins and outs of the fermentation lab and running bioreactors; it was actually lab-wise one of the most fun things I have learned during my PhD! Aside from being essential to the lab, you are also essential for the gezelligheid in IMB, thank you for all the fun dinners, movie nights, Sinterklaas, Friday drinks and trips to Belgium and Edinburgh! I am happy we are also friends outside of the lab, as bouldering and borreling buddy's! The countless strains, PCRs and hours of sequencing presented in this thesis would never have been possible without **Pilar,** who has taught me everything there is to know about molecular biology and sequencing. You are such a kind and fun person and am forever grateful for how you organized the lab. Also, **Agos**, it was short but sweet, but thank you for figuring out the mysteries of the sequencing lab together, and **Clara,** thank you for taking over all the molbio- and sequencing work after a time of difficulties! Talking about sequencing, thank you **Marcel**, for all of your help digging through the mountains of sequencing data I have generated, and for all the help organizing and analyzing it. I would never have thought I would use 'fun' and 'bioinformatics' in the same context, but I really enjoyed the collaboration we had on the RNAseq paper and enjoyed our time puzzling with the data!
Dear MSD-powervrouwen, **Jannie, Apilena, Gea** and **Astrid**, you are the silent motor behind every PhD project. I am especially grateful for your help with my bioreactor experiment, over almost 8 months you have autoclaved close to 1700 L (!) of medium for me and supplied me with approximately 50 bottles of ethanol without ever missing a beat, which I am super thankful for! Besides keeping the media flowing, you are always happy to think along, friendly, and interested in people, thank you for all the work you do! Apart from lab support, office support should also not be forgotten: **Michelle** and later **Susan,** thank you for keeping IMB turning and for puzzling with the PI-agendas to make life in IMB run smoothly, that must have been a challenge from time to time!

Mark, thank you for all the official and unofficial Mythical Mitochondria Meetings and guidance on the early stages of my project. You have taught me a lot, not just about mitochondria, respiration and physiology, and techniques, but also about the academic life and to have a critical attitude towards other, but more importantly, my own data (*Maar issie nou rood??*). I was happy to have you around in IMB, as a fellow JVN-fan but also as a friend and an example in how to pursue an academic career. Sadly, the pandemic and your new job got in the way of our joint mitochondria project, but I am super happy you have your own group now, we are colleagues once again, and I am grateful you wanted to join my committee, although I am bracing myself for your opposition…

Also to my other opponents, **Prof.dr. Liesbeth Veenhoff, Prof.dr. Martijn Huijnen, Prof.dr.ir. Stan Brouns and Prof.dr. John van der Oost**, thank you for critically reading my thesis and preparing an opposition, I am looking forward to discuss and debate the findings of my PhD with you!

During my PhD I had the great opportunity to collaborate with a lot of people outside of IMB, which I have learned a lot from. **Martin Pabst** and **Maxime**, thank you for teaching me the ins and outs of yeast proteomics and running experiments for me. I apologize for putting you on the impossible task of finding that one peptide, but I did learn a lot from it, it even inspired my choice for a postdoc! **Christophe Danelon,** and members of the **Danelon lab**, especially **David** and **Ana**, thank you for always welcoming me in your lab, it was really cool to experience working with *in vitro* systems and I really appreciated all the meetings and

discussions. Sadly, our collaboration did not make it in this thesis, but I am confident it will lead to a paper some day and hopefully will appear in the thesis of **Céline**, who I would like to thank for keeping the *in vivo* assembly alive after I could not go on with it. I am curious to see where it leads! I have also done many cool microscopy experiments at the Department of Bionanoscience, thank you to **Jérémie Capoulade, Anne Doerr** and **Michal Shemesh** for their advice and tireless help with microscopy experiments. Also a big thank you to all current and past members of the **BaSyC** consortium for the great meetings, summer schools and startup days, I have learned so much from all of you and it was great to be part of such a cool consortium!

It was an absolute pleasure to work together with so many talented and fun students during my thesis. Whether your work did or did not end up in this thesis, all of you have contributed so much to my PhD, not only in data, but also in fun time in the lab. Supervising each one of you was different and it taught me a lot about collaboration and teaching, and I hope you have enjoyed your theses as much as I did! **Gabriela**, you were the first to join my project and it was good to have someone on my side who knew as much about the project as I did, so we could figure it out together! It was good to see you back in iGEM, that means I must have not scared you away from lab work completely! **Angelique**, your enthusiasm and hard work have stuck with me, and the sole reason for not CRISPR-ing mitochondria was because I never got the RNA there, after your project we had the Cas9 and RT ready! **Kavicii,** you just took the project and ran with it, delivered an impressive thesis and introduced a number of new techniques to the project: FISH, Western blot and RNA-detection by fluorescence, that in the end are the core of this thesis. Next to that, you were an absolute blast to work with and enjoyed all the hours we spent in the dark microscopy lab with the fluorescent mitochondria as party lights! **Kirsten**, your suffered hardest from the pandemic, as you had to carry out the majority from home behind your computer. Nevertheless, you have finished it despite not being able to come to the lab and pipette anything, and I am proud of your hard work and resilience! **Valérie**, you also couldn't carry out the last most crucial experiments, but regardless of the fact you had to finish the thesis from home, I have seen you grow a lot, really enjoyed working in the CRISPEY team with you and Anna, and thank you again for a great pasta della norma recipe! **Askar**, despite not doing the project you came for and only being able to use the lab for 60 %, you single-handedly set up a whole new and challenging research line. Whether it was developing protocols or bioinformatics pipelines, your hard work and meticulous documentation have lead to a manuscript and I hope that by the time you read this, our joint efforts will even be published! **Quinte**, you took on maybe the most challenging research question in my project: a problem we didn't know how to solve and didn't even know how to screen for. We still don't know how to solve it, but thanks to you, your super hard work, millions of PCRs, digests and sequencing reactions, we at least know how to screen. I am sorry I wasn't there for the last part of your project but I am proud of how you persevered and finished with a great thesis! I am proud of all of you, and curious to see how your (academic) careers will develop!

Supervising a student is one challenge, supervising a team is a whole new level. Since 2016, iGEM has been a red thread in my scientific life. I would like to thank the **iGEM teams of 2019, 2020** for taking me along on their iGEM journeys as a supervisor, and it was an absolute pleasure to join the jamborees with all of you, and help you getting there. I am on top of that very happy to have been joined by an amazing team of **iGEM SUPERvisors (Ana, Martin, Elisa, David, Timon, Christophe, Esengul, Zoë and Britte),** with special mentions to **David F.** as a super fun yeast assembly-, Boston-, and judging buddy. We both joined the enemy lines now, and I hope to run into you at many Jamborees. **General Public Professor Britte**, in who I have not only found a rockstar-science friend (see you at the Nobel laureate dinner…) but also a great concert/festival friend, and **'Mama' Essie**, I will miss your gossip visits in my office and caring for our iGEM children.

I was happy that every other Friday I had a sounding board where I could discuss, troubleshoot and brainstorm about (my) research. **Paola, Francine, Eline, Sofia, Ewout,** and **all guest PhDs and students** who were a part of the ERC/Synbio meetings during my PhD, thank you for your input on my research and for taking me along on your glycolytic adventures! **Eline**, it was great being your paranymph together with **Ewout**!

The PhD was not only a time of cool experiments, but also a time of great friendships, parties, consumptions of yeast-fermented products, which resulted in a bunch of wonderful memories. **Arthur**, thanks to all your trust in me as a MEP-student I felt confident doing a PhD was the way to go, and I'm happy ever since!

You have been a mentor for my early PhD-days, and a friend for the rest. I promised that I would make it in academia after you left, so hope I'm still doing you proud! **Jonna**, it seems like we are destined to always work within 500 m of each other. I was really happy that we could work together for most of our PhDs, and that we are not only friends, but even became co-authors is the icing on the cake! You are one of the hardest-working and most creative people I know, and made IMB a fun and happy place to be at and I am happy we can continue to work in the same area for a little longer! **Marìa,** after iGEM it was clear that we could travel well together! It was great exploring Kenya with you, but also Gent, as long as there are boulder walls, cool adventures to be made or beers to try, we will have a good time! I promise this is the last thing I'm writing – after that we can travel again! **Jasmijn** and **Anna,** thank you for all the nice dinners and parties, and always bringing the fun and party to the group. **Aafke,** I really enjoyed sharing our big private office with you! I was happy to have someone with whom I could go through the hard parts of the PhD together, which mostly involved writing the thesis and trying to teach chemistry students what a microbe/bioreactor is… Thank you for all the laughs and hard work! **Sam,** thanks for all the coffees, beers and sharing our academic struggles, **Sophie** and **Susan**, I am really happy we went to Creative Life together, since it unleashed the artist in all of us! Thanks for organizing all the fun art and Bob Ross-nights. **Nicole**, you are an amazing asset to the group, and even though we never collaborated we somehow worked together a lot and I really enjoyed having you in the group. You have a great talent for science but are also a lot of fun, which is a precious combination! I really enjoyed our trip to Edinburgh and I hope we stay in touch!
Then some people who did not play a direct role in my PhD but still deserve a thank you: **Rinke**, for paving the way for young, female PIs, **Robert,** for being the most friendly, knowledgeable and enthusiastic-about-yeast person, with a second place and honorable mention for **Xavier**, who will always find time to brainstorm about complex yeast-puzzles with you, **Hannes,** thank you for teaching me the art of mito-isolation, **Marcel VL,** who will keep the mito-legacy alive (I can't wait to see what cool science you are going to do!), **Christiaan**, for keeping the reactors running alongside Erik, **Marieke** and **Rozanne,** who I trust will continue my legacy of playing indigestible rock and metal in the lab, **Jordi**, we go way back and it was cool that somehow we ended up colleagues, I had really fun times with you in the lab, **Sanne,** for all the serious talks on science-life, but also the uncontrollable laughter and lab-songs.
Although work-life balance is important, I can't help but thinking of IMB as a sort of 'home', which was due to the great, open and welcoming atmosphere. Thank you to **all of IMB** for the great coffee-table conversations, workdiscussions, brainstorming sessions, borrels, parties, conferences, late nights in the lab, Zoom-coffee breaks and more.

Gelukkig heb ik ook buiten het lab een groep fijne personen om me heen die me in alles gesteund hebben. **Lieke** en **Maxime**, bedankt voor alle wekelijkse etentjes, spelletjesavonden en foodie-weekends. Jullie hebben me vaak genoeg het lab uit moeten sleuren, maar ik ben blij dat jullie me er altijd aan herinnerd hebben dat er ook een leven is buiten gist! We wonen niet meer allemaal in Delft, maar ik hoop dat we door brouw-sessies en steeds ingewikkeldere spellen toch elkaar in ieder geval maandelijks blijven zien! **Djoeka**, bedankt voor alle puppy-therapy, dat was soms hard nodig. **Hudito Dames 7**, jullie zijn een fijn team dat me twee keer per week mijn werk liet vergeten en bedankt dat ik alle PhD-frustraties kwijt kon op het veld!

Ik heb een super lieve familie die me in alles steunt: **Papa** en **Mama**, en **Gerard** en **Patricia**, bedankt voor al jullie onvoorwaardelijke steun, hulp en liefde, maaltijden, verhuishulp, gezelligheid en een fijn warm nest. De afgelopen paar jaren zijn soms best zwaar geweest en het was fijn om altijd een fijn gezin te hebben waar ik op terug kon vallen. Bedankt voor altijd proberen te begrijpen wat ik doe, voor alle krantenknipsels over CRISPR. Het is de inspiratie geweest voor al het werk dat ik in wetenschapscommunicatie steek, en ik vond het altijd fijn om jullie trots op de eerste rij te zien zitten! **Olivia**, ik kan niet wachten tot je oud genoeg bent om mee te nemen naar NEMO en van jou een kleine wetenschapper te maken! Lieve **Oma**, je kan doctor Charlotje-potje niet meer meemaken, maar ik draag je altijd mee en je bent voor mij een voorbeeld van hoe sterk je als vrouw kan zijn.

Emiel, I don't wanna know what would've happened if I never had found your love […] Everything about you, I 👍. Je bent een stukje rust in alle chaos en ik weet niet hoe ik uiteindelijk mijn PhD had afgerond als wij niet midden in Covid verdwaald waren in de duinen. Bedankt voor al je zorgzaamheid, liefde en bovenal geduld voor me en ik heb ontzettend veel zin in alle avonturen die we gaan beleven in Utrecht!

CURRICULUM VITAE

Charlotte Koster was born on March 6th, 1993, in Voorburg, the Netherlands and lived in Rijswijk, Zuid-Holland, for the first 19 years of her life. She attended pre-university education (VWO-gymnasium) with a bilingual programme and obtained an additional International Baccalaureate (IB) in English Language and Literature. Upon graduation, Charlotte followed the bachelor's programme Life Science and Technology at Delft University of Technology and Leiden University. In her third year, she paused her studies to become a full-time board member of Study Association LIFE in the position of president, where she coordinated extracurricular and social events for Life Science & Technology students, including the organization of an international symposium on the molecular basis of aging and a study trip to Copenhagen and Berlin. Charlotte completed her bachelor's degree in 2015 with a minor programme in Advanced Life Sciences at Delft University of Technology and Leiden University and a Bachelor thesis internship at the Industrial Microbiology section of the department of Biotechnology of TU Delft, under the supervision of dr. Jasmine Bracher and Prof.dr. Ton van Maris on biotin metabolism in yeast.

Charlotte pursued a master's programme in Life Science & Technology at Delft University of Technology, with specialization track 'Cell Factory', and electives in numerical methods, modelling, and systems biology. As part of her master's degree, Charlotte joined the 2016 TU Delft iGEM team, with whom she aimed to develop improved fluorescence signals in bacteria and was awarded with three awards at the international Jamboree in Boston, MA. She performed a master thesis internship at the Industrial Microbiology section under supervision of dr. Arthur Gorter de Vries and Prof.dr. Jean-Marc Daran, where she developed a fluorescence-based GM-free method for industrial strain improvement. Upon completion of her thesis, Charlotte travelled to the United States for a 6-month internship at Ginkgo Bioworks in Boston, MA. She obtained her Master of Science (MSc.) degree in 2018, with the distinction *cum laude* and was awarded 'Best Graduate' 2018 of the Faculty of Applied Sciences at TU Delft.

In 2018, Charlotte moved back to the Netherlands and joined the Industrial Microbiology section of Delft University of Technology as a PhD candidate under supervision of Prof.dr.ir. Pascale Daran-Lapujade where she studied the potential of yeast homologous recombination for synthetic genomics and researched the possibility to genetically engineer the yeast mitochondrial genome, of which the work is described in the present thesis. Since March 2023, Charlotte has joined the Bacterial Genetics group of the Laboratory of Microbiology at Wageningen University & Research as a post-doctoral researcher under supervision of dr. Nico Claassens and Prof.dr. John van der Oost.

LIST OF PUBLICATIONS

Koster, C. C., Kohabir, K., den Ridder, M., Luttik, M., de Hulster, A.F.H., Pabst, M., Daran-Lapujade, P. (2023) Exploration of mRNA-sized RNA import into *Saccharomyces cerevisiae* mitochondria by a combined synthetic biology and adaptive laboratory evolution approach

Manuscript in preparation for submission.

Koster, C. C. & Daran-Lapujade, P. (2023). Arginine auxotrophy caused by deletion of the mitochondrial selectable marker *ARG8* in *Saccharomyces cerevisiae* can be reverted by loss-of-function mutations in *UME6*.

Manuscript under review for publication.

Koster, C. C., Kleefeldt, A., van den Broek, M., Luttik, M., Daran, J. M. G., & Daran-Lapujade, P. (2023). Long-read direct RNA sequencing of the mitochondrial transcriptome of *Saccharomyces cerevisiae* reveals condition-dependent intron abundance. *bioRxiv*, *2023-01 https://doi.org/10.1101/2023.01.19.524680*

Pre-print, *manuscript accepted for publication in Yeast*

Koster, C. C., Postma, E. D., Knibbe, E., Cleij, C., & Daran-Lapujade, P. (2022). Synthetic genomics from a yeast perspective. *Frontiers in Bioengineering and Biotechnology*, *10. https://doi.org/10.3389/fbioe.2022.869486*

Bouwknegt, J., **Koster, C. C.**, Vos, A. M., Ortiz-Merino, R. A., Wassink, M., Luttik, M. A., van den Broek, M., Hagedoorn, P.L., & Pronk, J. T. (2021). Class-II dihydroorotate dehydrogenases from three phylogenetically distant fungi support anaerobic pyrimidine biosynthesis. *Fungal biology and biotechnology*, *8*, *1-18. https://doi.org/10.1186/s40694-021-00123-6*

Gorter de Vries, A. R., **Koster, C. C.**, Weening, S. M., Luttik, M. A., Kuijpers, N. G., Geertman, J. M. A., Pronk, J. T., & Daran, J. M. G. (2019). Phenotype-independent isolation of interspecies *Saccharomyces* hybrids by dual-dye fluorescent staining and fluorescence-activated cell sorting. *Frontiers in Microbiology*, *10*, *871. https://doi.org/10.3389/fmicb.2019.00871*

Bracher, J. M., de Hulster, E., **Koster, C. C.**, van den Broek, M., Daran, J. M. G., van Maris, A. J., & Pronk, J. T. (2017). Laboratory evolution of a biotin-requiring *Saccharomyces cerevisiae* strain for full biotin prototrophy and identification of causal mutations. *Applied and Environmental Microbiology*, *83(16)*, *e00892-17. https://doi.org/10.1128/AEM.00892-17*

PATENTS

De Vries, A. R. G., **Koster, C. C.,** Daran, J. M. G., Geertman, J. M., & Kuijpers, N. G. A. (2021). Identification of rare products of crossing organisms.
European Patent Application No EP3790965A1