Investigation and engineering of respiratory energy coupling in yeasts

Jürgens, H.

DOI
10.4233/uuid:320d7140-4fad-4ffa-bc9a-50fcc344ed0f

Publication date
2022

Document Version
Final published version

Citation (APA)
Jürgens, H. (2022). *Investigation and engineering of respiratory energy coupling in yeasts*. [Dissertation (TU Delft), Delft University of Technology]. https://doi.org/10.4233/uuid:320d7140-4fad-4ffa-bc9a-50fcc344ed0f

Investigation and engineering of respiratory energy coupling in yeasts

Hannes Jürgens

Investigation and engineering of respiratory energy coupling in yeasts

Dissertation

for the purpose of obtaining the degree of doctor
at Delft University of Technology,
by the authority of the Rector Magnificus Prof.dr.ir. T.H.J.J. van der Hagen,
chair of the Board for Doctorates,
to be defended publicly on
Friday 13 May 2022 at 15:00 o'clock

by

Hannes JÜRGENS

Master of Science in Biotechnology
Aachen University of Applied Sciences, Germany,
born in Achim, Germany

This dissertation has been approved by the promotor and copromotor.

Composition of the doctoral committee:

Rector Magnificus	chairperson
Prof.dr. J.T. Pronk	Delft University of Technology, promotor
Dr.ir. R. Mans	Delft University of Technology, copromotor

Independent members:

Prof.dr. E. Nevoigt	Jacobs University, Bremen, Germany
Prof.dr. B. Teusink	Free University, Amsterdam, The Netherlands
Prof.dr.ir. M.C.M. van Loosdrecht	Delft University of Technology
Dr.ir. R.J. van Tatenhove-Pel	Delft University of Technology
Dr.ir. L. Wu	DSM, Delft, The Netherlands

Reserve member:

Prof.dr.ir. J.G. Daran	Delft University of Technology

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. This work was performed within the BE-Basic R&D Program (https://www.be-basic.org/), which was granted a FES subsidy from the Dutch Ministry of Economic Affairs, Agriculture and Innovation (EL&I). Research was performed in BE-Basic Flagship 10 in collaboration with Amyris, Inc. (Emeryville, CA, USA) and DSM (Delft, The Netherlands).

Cover: Laura Wenzel
Layout: Hannes Jürgens
Printed by: Ipskamp Printing B.V.
ISBN: 978-94-6421-746-9

An electronic version of this thesis is available at http://repository.tudelft.nl.

Table of contents

Summary

Microorganisms, enhanced by genetic engineering, are important cell factories for conversion of renewable feedstocks into fuels, chemicals, biomaterials, nutraceuticals and drugs. Microbial synthesis of these products from chemically simple carbon substrates often requires a net input of free energy in the form of ATP, which is typically provided by respiration in aerobic bioprocesses. In such processes, oxidation of a fraction of the substrate with oxygen releases carbon dioxide and water as final products and provides the ATP that is needed for product formation as well as for maintenance requirements and cellular growth.

In aerobic bioprocesses, the poor solubility of oxygen in water imposes rate limitations. Improvement of oxygen transfer by measures such as aeration and mixing, as well as cooling to compensate for respiration-associated heat generation, are major contributors to the cost of aerobic processes in microbial biotechnology. Since oxygen- and substrate requirements for ATP provision are negatively correlated to the ATP yield on oxygen from respiratory metabolism, one approach to alleviate these disadvantages and at the same time improve product yields of aerobic processes is to increase the ATP yield of respiration. An important factor that affects this respiratory efficiency is the composition of the respiratory electron-transport chain, into which electrons released by substrate dissimilation can enter at different points. Many eukaryotic cell factories can express both the efficiently-coupled (proton-pumping) Complex-I NADH dehydrogenase and less efficient alternative NADH dehydrogenases as entry points for electrons. Due to their difference in coupling efficiency, the relative contribution of these NADH dehydrogenases strongly influences the overall efficiency of ATP generation. A better understanding of condition-dependent expression and physiological relevance of the various NADH dehydrogenases would enable a rational design of strains and process conditions for increased production rates and yields.

Strains currently used for industrial production have not been selected based on their efficiency in respiratory coupling. For example, the highly popular yeast *Saccharomyces cerevisiae* does not possess the efficient Complex-I NADH dehydrogenase. The research presented in this thesis therefore focuses on two closely related Crabtree-negative yeasts, *Ogataea parapolymorpha* and *Ogataea polymorpha*, which were previously named *Hansenula polymorpha* and *Pichia angusta*. Not only do these yeasts possess a Complex-I NADH dehydrogenase, but they also exhibit other industrially relevant properties such as thermotolerance and methylotrophy.

Investigating the role of respiratory efficiency in yeast physiology requires knowledge on cellular mechanisms and pathways, which can for example be gained by investigating the physiological effects of targeted gene knockouts. One way to achieve

such knockouts is to develop efficient and effective genetic engineering methods. To make the potential of the recently established CRISPR-Cas9 system available for use in *O. parapolymorpha* and *O. polymorpha*, a Cas9-based genome editing tool was developed and described in **Chapter 2**. Such tools are typically based on species-specific plasmids and expression cassettes for Cas9 and guide-RNA expression, thereby limiting their use to a single species. CRISPR-Cas9 tools that function in multiple yeast species could contribute to an acceleration of research on non-conventional yeasts, for most of which fewer genetic tools are available. Therefore, in the research described in **Chapter 2** a multi-species plasmid carrying a pangenomic origin of replication was constructed that carried constitutive expression cassettes for Cas9 and ribozyme-flanked gRNAs. Its functionality was tested in two *Ogataea* species and also two *Kluyveromyces* species by targeting their *ADE2* orthologs. In *O. polymorpha* and *O. parapolymorpha*, Ade$^-$ mutants were not observed directly after transformation. However, when transformed cells were incubated in liquid medium for several days, *ADE2* targeting efficiencies between 9% and 63% were achieved, providing a straightforward and marker-free method for gene disruption mediated by non-homologous end joining (NHEJ). This system was then used to obtain an *O. parapolymorpha ku80* mutant, deficient in NHEJ, which enabled the marker-free deletion of *ADE2* by homologous recombination (HR), albeit at low efficiencies (<1%). Expression of a polycistronic gRNA array, consisting of two linked ribozyme-flanked gRNAs, also enabled simultaneous disruption of *ADE2* and *YNR1* in *O. parapolymorpha*. When tested in *K. lactis* and *K. marxianus*, the developed CRISPR-Cas9 tool achieved near-perfect targeting of the *ADE2* locus (≥96%), with at least 24% of transformants integrating a co-transformed DNA fragment via HR. These results indicate that this multi-species CRISPR-Cas9 tool might be applicable to other *Saccharomycotina* yeasts.

Despite the utilization of *O. parapolymorpha* and *O. polymorpha* for large-scale recombinant-protein production and numerous academic studies on their methanol metabolism, the growth energetics of these yeasts in aerobic cultures have not been comprehensively evaluated. In **Chapter 3,** the physiology of two industrially relevant wild-type strains of *O. parapolymorpha* and *O. polymorpha* was investigated in synthetic, glucose-containing medium under aerobic conditions. First, shake-flask cultures were utilized to determine specific growth rates at temperatures ranging from 30 to 49°C, with cultures at 40°C exhibiting the highest specific growth rates. Subsequently, growth physiology was quantified in tightly controlled bioreactor-grown batch- and chemostat cultures ($D = 0.1\ h^{-1}$) at 30 and 40°C to establish a wild-type strain baseline physiology. Both *O. parapolymorpha* and *O. polymorpha* showed a fully respiratory metabolism under the tested conditions, and exhibited biomass yields of ~0.5 g biomass dry weight per g glucose in both batch- and chemostat cultures.

Exact experimental procedures and data contexts, such as highly controlled bioreactor setups and connected analytics, are often not adequately captured by materials and methods sections of academic journals, despite having a potentially large impact on the recorded research outcome. The complex but standardized experimental work performed in **Chapter 3** was therefore used to evaluate a software package with the goal of addressing this challenge. The software utilizes flow sheets for systematic definition of experimental procedures, acquires and analyses data, and supports automatic linking of the generated data to the used procedures. While initial definition of the numerous experimental procedures was time intensive, they could afterwards easily be combined into the more complex shake-flask and bioreactor cultivation procedures needed for quantitative studies of microbial physiology. Furthermore, easy export of the captured, annotated data in an accessible form for data analysis allowed the removal of bias by efficient (scripted) calculation of the physiological strain properties.

The use of Complex-I NADH dehydrogenase instead of alternative NADH dehydrogenase for respiratory oxidation of NADH is expected to increase the ATP yield on the energy substrate and oxygen. However, as demonstrated in **Chapter 3**, *O. parapolymorpha* and *O. polymorpha*, which both contain the genes required for synthesis of the Complex-I NADH dehydrogenase, exhibited the same biomass yields on glucose as yeasts such as *S. cerevisiae* that have lost Complex I. Therefore, the research described in **Chapter 4** investigated the physiological importance of Complex I in *O. parapolymorpha*. To this end, a Complex I-deficient mutant was constructed using the CRISPR-Cas9 tool described in **Chapter 2**. Then, the wild-type strain and the congenic Complex I-deficient strain of *O. parapolymorpha* were grown on glucose in aerobic bioreactors operated in batch and chemostat mode. In batch cultures, both strains exhibited the same fully respiratory metabolism, growth rate and biomass yield, indicating that under these conditions, Complex I does not (notably) contribute to respiratory oxidation of NADH in *O. parapolymorpha*. In glucose-limited chemostat cultures, both strains again grew with a fully respiratory metabolism, however the Complex I-deficient mutant exhibited a 16 and 30% lower biomass yield on glucose and oxygen, respectively. This observation is consistent with a considerably lower efficiency of respiration in the Complex I-deficient mutant and suggests NADH oxidation by alternative NADH dehydrogenase in this strain under glucose-limited conditions at a specific growth rate of 0.1 h^{-1}. **Chapter 4** also describes a characterization of both *O. parapolymorpha* strains in retentostat cultures that allowed for a physiological comparison down to specific growth rates of approximately 0.001 h^{-1}, corresponding to doubling times of nearly one month. Compared to faster growing chemostat cultures, both wild-type and Complex I-deficient *O. parapolymorpha* were found to be capable of significantly reducing their maintenance energy requirements under these conditions,

and thus exhibited a 'stringent response'-like behavior. Data from quantitative transcriptome and proteome analyses performed on the batch-, chemostat- and retentostat cultures used throughout **Chapter 4** were consistent with physiological observations and indicated condition-dependent expression patterns of Complex I and alternative NADH dehydrogenases.

The results obtained in **Chapter 4** indicated that *O. parapolymorpha* harbors a branched respiratory chain that is able to oxidize NADH both via proton-pumping Complex-I NADH dehydrogenase and via one or multiple alternative NADH dehydrogenases. Whether such an alternative dehydrogenase can replace the function of Complex I in the oxidation of mitochondrial NADH depends on whether it is functionally expressed and exhibits suitable properties for NADH oxidation in the mitochondrial matrix. **Chapter 5** investigates the physiological role, localization and substrate specificity of the alternative NADH dehydrogenases of *O. parapolymorpha*. To this end, strains with various combinatorial deletions in structural genes encoding Complex I and alternative NADH dehydrogenases were constructed by a combination of using the CRISPR-Cas9 tool developed in **Chapter 2** as well as classical marker-based methods. The physiology of wild-type and mutant *O. parapolymorpha* strains was then characterized in glucose-grown shake-flask and chemostat cultures, and oxygen-uptake experiments with isolated mitochondria were used to examine exact mitochondrial localization of the alternative NADH dehydrogenases. To assess substrate specificity of the NADH dehydrogenases, they were individually overexpressed in *Escherichia coli* membranes and examined *in vitro*. Findings show that at least one of the alternative NADH dehydrogenases of *O. parapolymorpha* oxidizes mitochondrial ('internal') NADH (suggested name: OpNdi1), providing an explanation for the respiratory phenotype of Complex I-deficient *O. parapolymorpha* observed in **Chapter 4**. Moreover, the data suggest that OpNdi1 and Complex I are exclusively used in the presence and absence of excess glucose substrate, respectively. The respiratory chain of *O. parapolymorpha* was also found to contains at least one alternative NADH dehydrogenase that can oxidize cytosolic NADH and NADPH. Furthermore, a fully respiratory phenotype of glucose-limited *O. parapolymorpha* lacking all alternative NADH dehydrogenases indicates that additional mechanisms for oxidation of cytosolic NADH are present. With the exception of OpNdi1, whose elimination by gene deletion induced a Crabtree-positive phenotype in the presence of excess glucose, *O. parapolymorpha* possesses a flexible respiratory chain that tolerates disruption of multiple NADH dehydrogenases without compromising fully respiratory metabolism.

Samenvatting

Microorganismen, verbeterd door genetische modificatie, zijn belangrijke celfabrieken voor omzetting van hernieuwbare grondstoffen naar brandstoffen, chemicaliën, biomaterialen, voedingsmiddelen en medicijnen. Voor de microbiële synthese van deze producten uit chemisch gezien eenvoudige substraten is vaak extra 'metabole energie' nodig in de vorm van ATP. Deze ATP wordt in het algemeen geleverd via ademhaling in aerobe processen, waarbij een gedeelte van het substraat wordt geoxideerd met zuurstof en waarbij water en koolstofdioxide vrijkomen als eindproducten. Deze ATP wordt vervolgens gebruikt voor de vorming van het product, alsmede voor groei en onderhoud van de microbiële cellen.

In deze aerobe processen zorgt de slechte oplosbaarheid van zuurstof in water voor limitaties in de snelheid waarmee het eindproduct kan worden gevormd en voor extra kosten voor beluchting, mengen en koelen van de reactoren. Deze nadelen kunnen worden verminderd door de ATP-opbrengst van microbiële ademhaling te verbeteren, omdat de hoeveelheid substraat en zuurstof voor ATP-levering hier direct aan zijn gekoppeld. Een bijkomend voordeel is dat een verhoogde efficiëntie van ademhaling ook de productopbrengst verhoogd. Een belangrijke factor die de efficiëntie van de ademhaling beïnvloedt is de samenstelling van de elektronentransportketen die verantwoordelijk is voor de overdracht van elektronen uit substraat naar zuurstof. In deze keten kunnen elektronen op verschillende plaatsen binnenkomen. Veel eukaryote cellen kunnen hiervoor zowel het efficiënte, proton-pompende Complex-I NADH-dehydrogenase als de minder efficiënte alternatieve NADH-dehydrogenases gebruiken. Vanwege hun verschil in efficiëntie heeft de relatieve bijdrage van deze NADH-dehydrogenases een grote invloed op de totale efficiëntie waarmee ATP kan worden gevormd. Een beter begrijp van de conditie-afhankelijke expressie en fysiologische relevantie van deze systemen kan bijdragen aan de ontwikkeling van giststammen en processen waarnee verhoogde productiesnelheden en productopbrengsten kunnen worden gerealiseerd.

De gisten die op dit moment worden gebruikt voor industriële processen zijn niet geselecteerd vanwege de efficiëntie van hun ademhaling. De populaire gist *Saccharomyces cerevisiae* heeft bijvoorbeeld niet de beschikking over het efficiënte Complex-I NADH- dehydrogenase. Het onderzoek in dit proefschrift richt zich daarom op de nauw-verwante Crabtree-negatieve gisten *Ogataea parapolymorpha* en *Ogataea polymorpha*, eerder bekend als *Hansenula polymorpha* en *Pichia angusta*. Deze gisten beschikken niet alleen over een Complex-I NADH-dehydrogenase, maar hebben ook andere industrieel relevante eigenschappen, zoals een tolerantie tegen hoge

temperaturen en de mogelijkheid om methanol te gebruiken als koolstof- en energiebron.

Om de rol van de ademhalingsefficiëntie in de fysiologie van gisten te onderzoeken is het vergaren van kennis over de cellulaire mechanismen en stofwisselingsroutes, bijvoorbeeld door het analyseren van de effecten van 'knock-outs' van relevante genen, essentieel. Hiervoor zijn efficiënte methoden nodig om deze gisten te kunnen modificeren. Om het enorme potentieel van het recent ontdekte CRISPR-Cas9 systeem te kunnen aanwenden in *O. parapolymorpha* en *O. polymorpha* is, in het onderzoek dat in **Hoofdstuk 2** van dit proefschrift wordt beschreven, een op Cas9 gebaseerde methode ontwikkeld voor het maken van genetische modificaties in deze gisten. Het is gebruikelijk om dergelijke methoden te ontwikkelen door gebruik te maken van soort-afhankelijke plasmiden en cassettes voor de expressie van Cas9 en het guide-RNA. Hiermee wordt echter het gebruik van de methoden gelimiteerd tot een enkele soort. CRISPR-Cas9 methoden die werken in meerdere gistsoorten kunnen bijdragen aan het versnellen van het onderzoek aan niet-conventionele gisten, waarvoor vaak minder methoden voor genetische modificatie ontwikkeld zijn. Om deze reden werd in het in **Hoofdstuk 2** beschreven onderzoek een plasmide ontwikkeld dat werkt in meerdere soorten. Hierbij werd gebruik gemaakt van een pangenomisch startpunt voor DNA-replicatie en van constitutieve expressiecassettes voor Cas9 en ribozym-geflankeerde gRNAs. De functionaliteit van dit plasmide werd daarna getest in twee *Ogataea* en twee *Kluyveromyces* soorten, waarbij orthologen van het *ADE2* gen als doelwit werden gebruikt. In *O. polymorpha* en *O. parapolymorpha* werden geen Ade$^-$ mutanten gevonden direct na transformatie. Echter, nadat de cellen meerdere dagen in vloeibaar medium waren geïncubeerd werden efficiënties van 9% en 63% voor mutaties in *ADE2* waargenomen. Hiermee kan deze methode direct worden toegepast voor het muteren van genen, zonder gebruik van genetische markers, via 'non-homologous end joining' (NHEJ). Dit systeem werd vervolgens gebruikt in een *ku80* mutant van *O. parapolymorpha*, die niet meer in staat is om NHEJ uit te voeren. In deze stam was het mogelijk om via homologe recombinatie (HR) het *ADE2* gen te verwijderen, al was de efficiëntie laag (<1%). Het tot expressie brengen van een 'polycistronic gRNA array', bestaande uit twee verbonden, ribozym-geflankeerde gRNAs, maakte het bovendien mogelijk om tegelijkertijd mutaties te maken in *ADE2* en *YNR1* in *O. parapolymorpha*. In *K. marxianus* en *K. lactis* was het met de nieuw ontwikkelde CRISPR-Cas9 methode mogelijk om met zeer hoge efficiëntie (≥96%) het *ADE2* locus te muteren, waarbij ten minste 24% van de transformanten in staat was om via HR een getransformeerd DNA-fragment te integreren. Deze resultaten toonden aan dat deze CRISPR-Cas9 methode mogelijk ook in andere *Saccharomycotina* gisten kan worden gebruikt.

Ondanks het feit dat *O. parapolymorpha* en *O. polymorpha* gebruikt worden voor het op industriële schaal produceren van recombinante eiwitten en er veel studies zijn gedaan naar hun methanolmetabolisme, is de energetica van aerobe groei van deze gisten niet uitvoerig bestudeerd. In **Hoofdstuk 3** werd daarom de groei van twee industrieel relevante, wilde stammen van beide gisten onderzocht in synthetisch medium, in de aanwezigheid van glucose en zuurstof. Eerst werd in schudkolven onderzocht wat de specifieke groeisnelheid van de gisten was bij temperaturen van 30 tot 49°C. In deze experimenten werd de snelste groei waargenomen bij 40°C. Vervolgens werden de groeikarakteristieken bepaald in nauwkeurig gecontroleerde bioreactoren die als batchcultuur of als chemostaatcultuur ($D = 0.1\ h^{-1}$) werden bedreven bij temperaturen van 30 en 40°C. Beide gisten vertoonden een volledig respiratoir metabolisme onder de geteste condities, waarbij de biomassaopbrengst in zowel batch- als chemostaatcultures ongeveer 0.5 gram droge biomassa per gram glucose bedroeg.

In de wetenschappelijke literatuur worden de experimentele procedures en de context van de verworven data, zoals in complexe bioreactoropstellingen, vaak niet volledig beschreven in de het hoofdstuk “Methods”. Deze factoren kunnen echter een grote invloed kunnen hebben op de resultaten van het beschreven werk. De gestandaardiseerde maar complexe experimenten die beschreven zijn in **Hoofdstuk 3** werden daarom gebruikt om te testen of een softwarepakket deze uitdaging op kan lossen. Deze software gebruikt ‘flow sheets’ om experimentele procedures systematische te beschrijven en koppelt meetresultaten automatisch aan deze procedures. Hoewel het definiëren van de verschillende experimentele deelprocedures in eerste instantie veel tijd kostte, bleek het vervolgens eenvoudig om deze samen te voegen en zo complexere experimentele procedures voor fysiologische studies in schudkolven en bioreactoren te beschrijven. Een bijkomend voordeel hiervan was dat de experimentele metingen naderhand eenvoudig konden worden geëxporteerd en vervolgens objectief door automatische scripts kon worden doorgerekend.

Gebruik van een Complex-I NADH dehydrogenase voor oxidatie van NADH via ademhaling, in plaats van de alternatieve NADH dehydrogenases, zorgt naar verwachting voor een hogere ATP-opbrengst op zuurstof en op het energiesubstraat. Echter, zoals beschreven in **Hoodstuk 3** was de biomassa-opbrengst van *O. parapolymorpha* en *O. polymorpha*, die allebei genen benodigd voor het maken van Complex-I bevatten, gelijk aan die van gisten zoals *S. cerevisiae* waarin geen Complex I voorkomt. Om deze waarneming nader te onderzoeken richtte **Hoofdstuk 4** zich daarom op de fysiologische rol van Complex I in *O. parapolymorpha*. Hiertoe werd eerst met de in **Hoofdstuk 2** beschreven CRISPR-Cas9 methode een mutant gemaakt waarin Complex I niet meer actief is. Vervolgens werd groeien van de mutant en het wildtype in aerobe batch- en chemostaatcultures vergeleken. In aerobe batchcultures was groei

van beide stammen volledig respiratoir, en waren de groeisnelheid en biomassaopbrengst gelijk. Dit resultaat wees erop dat Complex I onder deze kweekcondities geen prominente rol speelt in de oxidatie van NADH. Hoewel ook in de glucose-gelimiteerde chemostaatcultures de beide stammen volledig respiratoir groeiden, vertoonde de Complex I-mutant onder deze omstandigheden een 16 en 30% lagere biomassaopbrengst op respectievelijk glucose en zuurstof. Deze waarneming sloot goed aan bij het verwachte verschil in efficiëntie wanneer de mutant in plaats van Complex I de alternatieve NADH dehydrogenases gebruikt. In **Hoofstuk 4** werd de Complex I-mutant ook vergeleken met het wildtype in retentostaatcultures, waarin fysiologische karakterisering bij zeer lage groeisnelheden van 0.001 h^{-1} (verdubbelingstijd van bijna een maand) mogelijk was. Onder deze omstandigheden bleken beide stammen in staat de energiebehoeften voor celonderhoud ('maintenance') significant te verlagen ten opzichte sneller groeiende cellen. Dit verschijnsel lijkt sterk op de 'stringent response' die in andere microorganismen is waargenomen. Transcriptoom- en proteoomdata uit de batch-, chemostaat- en retentostaatculturen die zijn beschreven in **Hoofdstuk 4** waren in overeenstemming met de fysiologische waarnemingen en wezen eveneens op een conditie-afhankelijke expressie van Complex I en de alternatieve NADH-dehydrogenases.

De resultaten uit **Hoofdstuk 4** leken erop te wijzen dat *O. parapolymorpha* een vertakte ademhalingsketen heeft waarmee NADH zowel kan worden geoxideerd via het proton-pompende Complex I als via een of meer alternatieve NADH dehydrogenases. Of alternatieve NADH dehydrogenases in staat zijn om de functie van Complex I over te nemen voor het oxideren van mitochondrieel NADH, is zowel afhankelijk van hun expressie als van de mogelijkheid om NADH die wordt gevormd in de mitochondriële matrix te kunnen oxideren. **Hoofdstuk 5** beschrijft daarom onderzoek naar de fysiologische rol, lokalisatie en substraatspecificiteit van deze alternatieve NADH dehydrogenases in *O. parapolymorpha*. Hiervoor werden eerst stammen gemaakt waarin Complex I en de alternatieve NADH dehydrogenases in verschillende combinaties werden uitgeschakeld, daarbij gebruik makend van de in **Hoofdstuk 2** beschreven methode en van klassieke methoden. Vervolgens werd de fysiologie van de resulterende stammen bepaald bij groei op glucose in schuldkolven en chemostaatcultures. Tevens werden zuurstofopnameexperimenten met geïsoleerde mitochondriën uitgevoerd om de locatie van de alternatieve NADH dehydrogenases te bepalen. Om de substraatspecificiteit van deze enzymen te bepalen werden ze bovendien tot overexpressie gebracht in *Escherichia coli*-membranen en daarna *in vitro* bestudeerd. Deze experimenten toonden aan dat ten minste een van de alternatieve NADH dehydrogenases in *O. parapolymorpha* mitochondrieel ("intern") NADH kan oxideren. Het betrokken eiwit werd OpNdi1 genoemd. Deze waarneming verklaarde het respiratoire fenotype van de Complex I-deficiënte mutant dat is beschreven in

Hoofdstuk 4. Deze data wijzen er bovendien op dat OpNdi1 en Complex I exclusief gebruikt worden in respectievelijk de aan- en afwezigheid van hoge glucoseconcentraties in het medium. Daarnaast werd ontdekt dat de ademhalingsketen van *O. parapolymorpha* tenminste één alternatief NADH- dehydrogenase bevat dat zowel cytosolisch NADH als NADPH kan oxideren. Tot slot wees het vermogen van een mutant waarin alle alternatieve NADH dehydrogenases waren uitgeschakeld, desondanks volledig respiratoir te groeien, erop dat er in *O. parapolymorpha* additionele, nog onbekende mechanismen zijn om cytosolisch NADH te oxideren. Met uitzondering van OpNdi1, waarvan de deletie zorgt voor een Crabtree-positief fenotype bij hoge glucoseconcentraties, heeft *O. parapolymorpha* een flexibele ademhalingsketen, waarin meerdere NADH dehydrogenases kunnen worden uitgeschakeld, zonder dat dit het vermogen om volledig respiratoir te kunnen groeien aantast.

Chapter 1

General introduction

Definition of biotechnology. The term biotechnology was first coined in 1919 in Berlin, where Hungarian agricultural engineer Károly Ereky published his book entitled "Biotechnologie der Fleisch-, Fett- und Milcherzeugung im landwirtschaftlichen Großbetriebe für naturwissenschaftlich gebildete Landwirte" (Biotechnology of meat, fat and milk production in an agricultural large-scale farm). Under the term 'biotechnology', Ereky included all procedures that turned feedstocks into consumer goods with the use of living organisms, mentioning as examples a beet that turns the carbonic acid from the air (*Kohlensäure der Luft*) into sugar, and the cow that turns fodder into milk (Ereky 1919, VDI 2015). Nowadays, biotechnology is deliberately defined much broader, for example by the OECD: "the application of science and technology to living organisms as well as parts, products and models thereof, to alter living or non-living materials for the production of knowledge, goods and services" (OECD 2013). This definition encompasses modern applications, such as microbial, cell and tissue cultures, applications of DNA, RNA and protein molecules, as well as bioinformatics and bioprocess technology development.

Modern industrial biotechnology produces a wide range of products, including bulk and fine chemicals, transport fuels, (bio)materials, nutraceuticals, proteins and medicinal drugs (Meyer *et al.* 2016). A key option that industrial biotechnology offers in comparison with fossil-based production of these goods is environmental sustainability in the form of reducing carbon dioxide emissions. Substrates utilized for biotechnological production are typically plant-based sugars or sugar-containing streams, which are themselves formed from carbon dioxide by green plants. With humanity rapidly exhausting the carbon budgets predicted to limit the anthropogenic rise in global temperatures to manageable levels (Rogelj *et al.* 2018), there is a constant drive to replace fossil resource-based productions and further improve the efficiency of biotechnological processes.

History of yeast industrial biotechnology. The earliest production of wine and beer by humans can now be traced back some 6,000 and 13,000 years, respectively (Barnard *et al.* 2011, Liu *et al.* 2018), and products of alcoholic fermentation were likely already consumed by the ancestors of modern humans (Dudley 2004). However, it was not until the 1830s that researchers were independently able to demonstrate that alcoholic fermentation is linked to living yeast, observed as globular bodies (*globules*, *Kügelchen* or *Körnchen*) capable of reproduction (Barnett 1998, Cagniard-Latour 1838, Kützing 1837, Schwann 1837). Based on his experiments, Theodor Schwann, a physiologist from Berlin, concluded that sugar is a growth substrate for yeast and referred to it as *Zuckerpilz*, 'sugar-fungus', which led to the introduction of the genus name *Saccharomyces* (Meyen 1838, Schwann 1837). Around the time of these discoveries, production of fermented beverages was already occurring at significant scales. For

example, in France, 35 million hectoliters of wine were produced in 1823 (Barnett 1998), while beer production in Germany occurred at an estimated 23 million hectoliters in 1840 (Buchholz and Collins 2013, Knapp 1847).

Initially motivated by production inconsistencies, Louis Pasteur regularly visited an alcohol factory during the 1850s to take samples from the fermentation broth. He observed that budding yeast was present in normal fermentation runs but found rod-shaped 'lactic acid yeast' when the process failed. This observation led Pasteur to delve more deeply into the investigation of fermentation, and as a result he described and demonstrated several concepts and procedures that would form the basis of modern industrial biotechnology. These incuded association of different fermentation processes with different specific microbes, the practice of isolating pure cultures and using inoculation to increase process consistency, and the notion that nutrients for the culture must be provided by the fermentation medium (Barnett 2000, Pasteur 1858a, Pasteur 1858b). Based on this work Pasteur later also developed a closed fermentation vessel that would protect the process from contaminants (Pasteur 1873).

During the remainder of the nineteenth century, the yeast fermentation industry kept growing, manufacturing mainly increasing amounts of beer and wine as well as ethanol and pure yeast for industrial (food) applications (Buchholz and Collins 2013, Ullmann 1915). In the early- to mid- twentieth century, yeast processes saw technological advances such as the development of the fed-batch process (*Zulaufverfahren*) for efficient production of *Saccharomyces cerevisiae*, also known as baker's yeast (Intern.YeastCo.Ltd. 1933). Other, novel bioprocesses developed during this time were largely based on bacteria and fungi producing, for example, butanol, acetone and glycerol, citric acid and penicillin, as well as enzymes from various natural producers (Buchholz and Collins 2013, Bud 2007, Porro *et al.* 2011).

Over the last 50 years, biotechnology saw the development of genetic engineering and recombinant DNA (rDNA) technology (see below). The demand for various products that could be synthesized in organisms such as yeast, combined with the development of metabolic engineering and recombinant protein production led to a range of new yeast-based industrial bioprocesses. These processes are aimed for example at the production of metabolites that can be used to replace polymers produced from non-renewable resources, such as such as lactic acid and succinic acid (Es *et al.* 2018, Sauer *et al.* 2010, Saxena *et al.* 2017), as well as recombinant therapeutic proteins such as insulin (Baeshen *et al.* 2014, Ferrer-Miralles *et al.* 2009, Mattanovich *et al.* 2014). Genetic engineering also is a key tool in establishing second generation bioethanol production aimed at reducing the carbon footprint of biofuel production and its competition with the food industry, by making carbon sources accessible that established yeast biofuel production hosts (i.e., *S. cerevisiae*) cannot naturally utilize (Carus and Dammer 2013, Jansen *et al.* 2017).

Besides the well-established production host *S. cerevisiae*, so-called 'non-conventional' yeasts are becoming increasingly popular in industrial biotechnology applications due to their promising properties (Forti *et al.* 2015, Radecka *et al.* 2015). For example, *Scheffersomyces stipitis* (*Pichia stipitis*) can naturally utilize xylose, a major constituent of lignocellulose found in agricultural waste streams (Toivola *et al.* 1984), and methylotrophic *Pichia pastoris (Komagataella pastoris*) has proven to be highly suitable for the production of many recombinant proteins (Mattanovich *et al.* 2012). Another interesting non-conventional yeast is the methylotroph *Hansenula polymorpha* (now *Ogataea polymorpha* and *Ogataea parapolymorpha*), which, in addition to native utilization of lignocellulose constituents and suitability for (large-scale) recombinant protein production, exhibits thermotolerance up to 50°C (Mattanovich *et al.* 2012, Ryabova *et al.* 2003).

Substrate dissimilation provides cellular energy for growth, maintenance and (non-dissimilatory) product formation. Yeasts and other microbes used as the workhorses of industrial biotechnology ideally function purely as a catalyst that allows fast and complete conversion of the utilized substrate into the desired product. However, in a microbial bioprocess in which cells grow and thereby form new biomass, the substrate cannot be fully utilized for product formation. In growing cultures, substrate is required for cell proliferation, as it is assimilated into new biomass including the synthesis of the enzymes catalyzing the desired reaction steps for product formation. In addition, a portion of the substrate has to be catabolized (= dissimilated) to provide cellular energy in the form of adenosine triphosphate, ATP, the 'energy currency' of cells. ATP is required for the formation of new biomass from substrate and for cellular maintenance, which includes for example the turnover of macromolecular components and the maintenance of concentration gradients across biological membranes (Pirt 1965, van Bodegom 2007). Substrate dissimilation for ATP formation can occur by any dissimilatory (net ATP yielding) pathway. For example, in most yeast species glucose can either be converted via alcoholic fermentation, which forms ethanol and carbon dioxide, or via aerobic respiration which results in carbon dioxide and water. Cells typically balance their ATP production with their ATP requirements and, with some exceptions (Russell 2007), avoid unnecessary formation and 'spilling' of ATP in order to maximize proliferation (new biomass) and fitness (survival/maintenance). For growing cells, distribution of substrate over biomass formation and ATP synthesis can be illustrated as shown in **Figure 1.1A**.

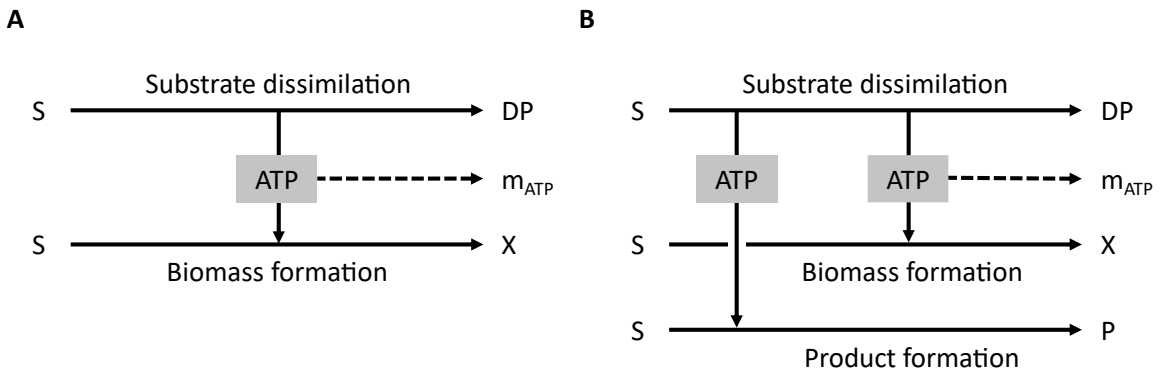


Figure 1.1: The distribution of substrate utilization in growing cells. **A**: Growing cells have to dissimilate a certain amount of substrate (S) in order to provide ATP for formation of biomass (X) and maintenance energy requirements (m_{ATP}). Substrate dissimilation results in the formation of a dissimilatory product (DP). **B**: If a desired product (P) requires the net-input of ATP, substrate dissimilation needs to provide ATP for product formation in addition to biomass formation and maintenance energy requirements.

Products other than microbial biomass can be categorized based on the ATP stoichiometries of their synthesis pathways. Into which category a specific product molecule falls into is determined by the Gibbs energy of formation of the utilized substrate and the desired product, as well as the metabolic reaction or combination of metabolic reactions (i.e., the metabolic pathway) that connect the two. For dissimilatory products, the overall reaction occurs with a negative change in Gibbs free energy (ΔG), making it thermodynamically favorable to run in the desired direction (from substrate to product). Depending on the metabolic pathway, part of this free energy can be conserved in the form of ATP. The distribution of substrate and ATP when producing a dissimilatory product can again be illustrated by **Figure 1.1A**, with the product being the end result of substrate dissimilation. In a stark contrast to the generally assumed 'objectives' of a growing cell, bioprocess engineers aiming to produce a dissimilatory product will typically attempt to minimize the amount of substrate spent on cellular growth and maintenance in order to achieve a favorable production process.

In contrast to dissimilatory products, formation of a non-dissimilatory product requires a net input of ATP to drive reactions which would otherwise exhibit a positive change in Gibbs free energy. Non-dissimilatory products in bioprocesses are for example (recombinant) proteins such as enzymes and antibodies, but also smaller molecules or metabolites whose formation from substrate require the net-input of ATP. This situation occurs, for example, during the production of farnesene (Meadows *et al.* 2016) or medium-chain fatty acids (Zhu *et al.* 2020) from glucose. For non-dissimilatory products, the dissimilation of substrate must not only provide ATP for biomass formation and maintenance energy requirements, but also for formation of the non-dissimilatory product itself (**Figure 1.1B**).

Aerobic bioprocesses for non-dissimilatory products. Economic considerations are of great importance for biotechnological process and, besides volumetric productivity and titer, the product yield on substrate is a key factor in determining overall process economics. Because the formation of non-dissimilatory products requires an input of ATP, generating the ATP as efficiently as possible from the substrate can positively affect the yield of such a process. As aerobic respiration provides vastly more ATP from the same amount of carbohydrates than other, fermentative dissimilatory pathways (Madigan *et al.* 2014), processes for biosynthesis of non-dissimilatory products are typically performed aerobically, in the presence of oxygen, resulting in a smaller portion of substrate required for ATP synthesis.

One major downside of an aerobic bioprocesses is the requirement for dissolved molecular oxygen in the liquid phase. The solubility of oxygen in an aqueous solution is low, and therefore it constantly needs to be supplied and transferred from the gas phase to the liquid phase which results in additional process costs for aeration and mixing. The oxygen transfer rate (OTR) into the liquid phase can easily be surpassed by the oxygen uptake rate (OUR) at which cells take up the dissolved oxygen, especially at high cell densities and/or when the cultured cells exhibit high OURs (Doran 2013). If this is the case and dissolved oxygen in the liquid phase is depleted, aerobic respiration can no longer fully satisfy ATP requirements. As a result, cellular metabolism alters, leading either to (partial) substrate dissimilation via oxygen-independent pathways with a lower ATP yield or to growth arrest (Weusthuis *et al.* 1994). A special case is represented by the Crabtree effect, typically described for *Saccharomyces* yeasts but also present in other yeast genera and other types of cells. Cells that show the Crabtree effect exhibit aerobic fermentation, i.e. substrate dissimilation via a low-ATP oxygen-independent pathway, even when dissolved oxygen is available and therefore require additional strategies to maximize efficient ATP generation (Dashko *et al.* 2014, De Deken 1966). To prevent oxygen depletion in cultures with high OUR, aerobic bioprocesses are often monitored for the dissolved oxygen concentration and operated in fed-batch mode, which allows operators to control the specific growth rate (and thereby OUR) by gradual substrate addition (Sonnleitner and Chmiel 2011). As a result, volumetric productivity of industrial scale bioprocesses is usually limited by oxygen transfer from the gas to the liquid phase (Hensing *et al.* 1995).

Aerobic respiration (sugar dissimilation + oxidative phosphorylation) provides cellular energy in the form of ATP. If enough oxygen is available, the energy substrate, e.g. glucose, can be completely respired to provide cellular energy in the form of ATP. Respiratory dissimilation of organic substrates occurs via the combined action of dissimilatory oxidation of the substrate to carbon dioxide and oxidative phosphorylation by the respiratory chain. These separate processes are linked by the

NAD^+/NADH redox couple, which serves as the predominant biological carrier of reductive power and plays a paramount role in redox metabolism. Due to the small amounts of NAD^+/NADH present in the cell, the rate of NAD^+ reduction and NADH oxidation need to be constantly balanced (Xiao *et al.* 2018),

In the dissimilation of glucose ($C_6H_{12}O_6$) via the Embden–Meyerhof–Parnas (EMP) pathway found in yeasts, a molecule of glucose is first oxidized in a multi-step metabolic pathway, resulting in the formation of 2 pyruvate, 2 ATP from substrate-level phosphorylation (SLP) and 2 NADH. The 2 molecules of pyruvate are then further oxidized by the combined action of the pyruvate-dehydrogenase complex and the enzymes constituting the tricarboxylic acid (TCA) cycle, eventually resulting in the formation of 6 carbon dioxide molecules and yielding another 2 ATP equivalents from SLP and 10 NADH (equivalents) in the process. Then, the NADH molecules generated by glucose dissimilation are (re)oxidized by the respiratory chain via the process of oxidative phosphorylation, regenerating NAD^+ and producing additional ATP (**Figure 1.2**).

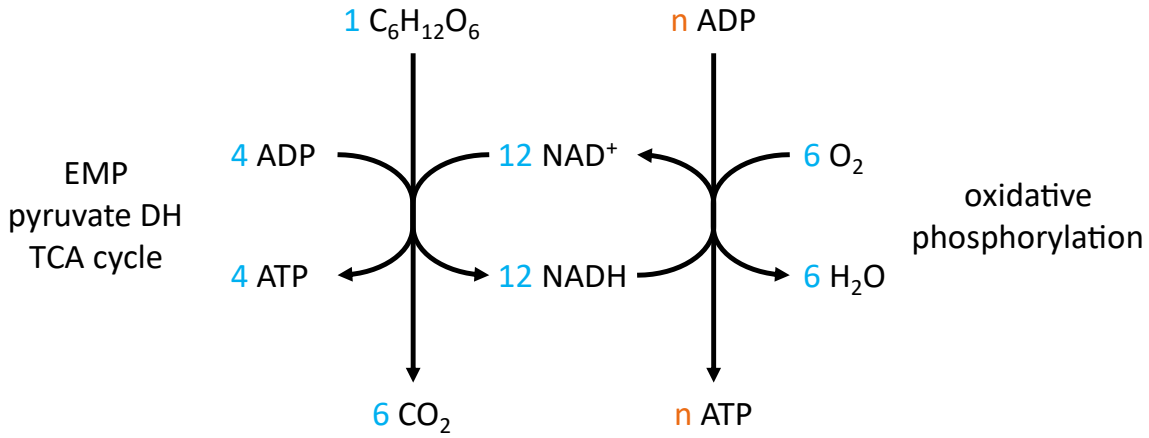


Figure 1.2: Aerobic respiration. The left side shows the oxidation of glucose ($C_6H_{12}O_6$) into carbon dioxide (CO_2) by the Embden–Meyerhof–Parnas (EMP) pathway, pyruvate dehydrogenase (pyruvate DH) and tricarboxylic acid cycle (TCA cycle), resulting in the generation of reducing equivalents (NADH) and ATP via substrate level phosphorylation (SLP). The right side shows (re)oxidation of the produced NADH by the respiratory chain in the process of oxidative phosphorylation, which yields additional ATP and consumes oxygen. For the sake of simplicity, the 2 molecules of $FADH_2$ generated by succinate dehydrogenase as part of the TCA cycle are depicted as NADH.

While the dissimilation of glucose via this pathway yields a fixed amount of ATP via SLP, the stoichiometry of ATP produced via oxidative phosphorylation from the 12 NADH, i.e. the efficiency of respiratory energy coupling, varies greatly between different species and conditions (approximately between 12 and 32) and depends on a range of factors such as compartmentation and respiratory chain composition (see below). Despite achieving less than the thermodynamically possible yield of ATP under

physiologically relevant conditions[1], aerobic respiration yields vastly more ATP per glucose than anaerobic pathways such as e.g. yeast alcoholic fermentation, which only generates 2 ATP per glucose (Villadsen *et al.* 2011).

Compartmentation and composition of aerobic respiration in eukaryotes. In eukaryotes such as yeasts, oxidation of glucose and the corresponding NADH formation occurs partly in the cytosol (EMP glycolysis) and partly in the mitochondrial matrix (pyruvate dehydrogenase and TCA cycle). Reoxidation of the formed NADH via oxidative phosphorylation is catalyzed by the respiratory chain and takes place within the mitochondria, more specifically at the inner mitochondrial membrane (IMM) (Gilkerson *et al.* 2003). As the IMM serves as a barrier for NADH/NAD^+ between the mitochondrial matrix and the surrounding intermembrane space/cytosol (von Jagow and Klingenberg 1970), redox metabolism is compartmentalized and NADH has to be reoxidized in the compartment in which it is generated. Therefore, eukaryotic respiratory chains must possess the means to directly oxidize NADH on both sides of the IMM and/or indirect mechanisms such as shuttles that allow transfer of NADH (equivalents) between the different compartments (Bakker *et al.* 2001).

Eukaryotic respiratory chains are comprised of a sequence of enzymes and protein complexes in the IMM and function according to chemiosmotic coupling. Electrons released by the oxidation of NADH are led along the sequence of protein complexes towards the electron acceptor oxygen. The free energy released by these reactions is coupled to the translocation of protons across the IMM, establishing a proton gradient which can in turn be utilized for ATP formation (Madigan *et al.* 2014, Mitchell 1966).

Respiratory chains have evolved to be highly diverse and versatile, and different species harbor different configurations of proteins and enzyme complexes within their respiratory chains (Marreiros *et al.* 2016a). In the canonical eukaryotic (animal) respiratory chain, NADH is oxidized exclusively via Complex I NADH dehydrogenase (also called respiratory Complex I, NADH:ubiquinone oxidoreductase or NDH1). The electrons released by this oxidation are then transported via ubiquinone to Complex III (cytochrome *bc1* complex), from which they are shuttled via cytochrome *c* to Complex IV (cytochrome *c* oxidase), where finally reduction of oxygen to water occurs. Respiratory complexes I, III and IV couple the passage of electrons and the accompanying redox reactions to net translocation of 4, 2 and 4 protons across the IMM, respectively. Finally, the F_0F_1 ATP synthase (Complex V, ATPase), also located in the

[1] At a change in free energy of 220 kJ for the oxidation of 1 mol NADH to NAD^+ and H_2O and 40-55 kJ/mol for the activation of 1 mol ADP to ATP under physiological conditions (depending on the exact conditions present; e.g., 54 kJ free energy must be spent at pH = 7, [ATP/ADP] = 13.4 and [Pi] = 1 mM), 4 – 5.5 ATP could be synthesized from oxidation of a single NADH, or at least 48 ATP from oxidation of 1 glucose (calculated as 12 NADH equivalents).

IMM, uses this proton gradient for the formation of ATP (Jones *et al.* 2017, Madigan *et al.* 2014) (**Figure 1.3**).

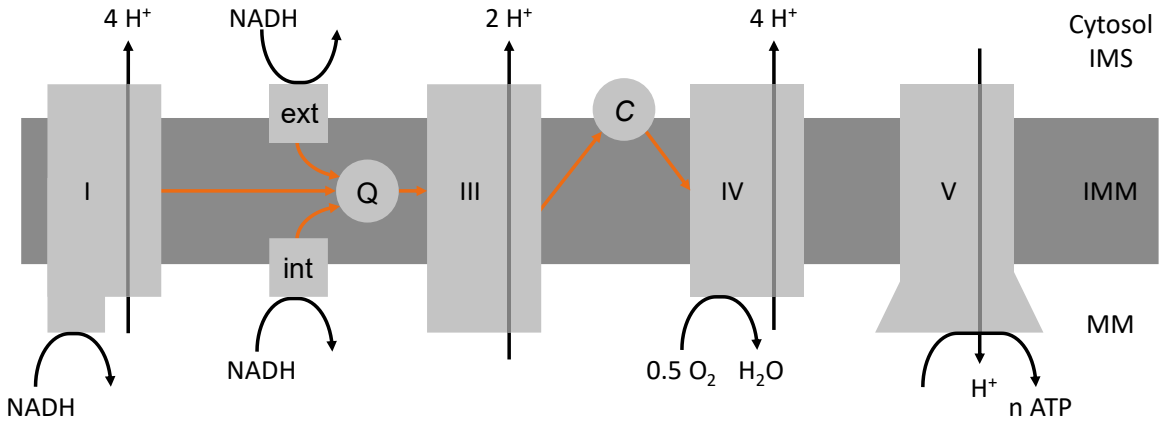


Figure 1.3: The mitochondrial respiratory chain in the inner mitochondrial membrane (IMM). Respiratory Complex I (I) and 'internal' alternative NADH dehydrogenase(s) (int) oxidize NADH in the mitochondrial matrix (MM). 'External' alternative NADH dehydrogenase(s) (ext) can directly oxidize cytosolic NADH. All NADH-oxidizing enzymes transfer electrons (orange arrows) released by the oxidation to the quinone pool (Q), from where they are linearly funneled through the remainder or the respiration chain, passing through Complex III (III), cytochrome *c* (*c*) and Complex IV (IV), before reducing oxygen to water. Complexes I, III and IV, but not the alternative dehydrogenases, contribute to the proton gradient across the IMM which is utilized by F0F1 ATPase (V) for formation of ATP. Note that neither Complex I nor the internal or external alternative dehydrogenases are essential or conserved. IMS: intermembrane space.

In contrast to the linear animal respiratory chain, the respiratory chains of yeasts are often branched and contain alternative routes for NADH oxidation and/or oxygen reduction (Joseph-Horne *et al.* 2001). Besides respiratory Complex I, so-called 'alternative NADH dehydrogenases' (type-II NADH dehydrogenases, NDH2) are common. These monotopic, single-subunit enzymes catalyze the same NADH:ubiquinone oxidoreduction reaction as Complex I, but do not translocate protons over the IMM in the process. In addition, depending on the exact localization of these proteins, their catalytic site can face either the mitochondrial matrix (localized to inner side of IMM) or the cytosol (localized to outer side of IMM), allowing direct oxidation of cytosolic NADH (**Figure 1.3**) (Antos-Krzeminska and Jarmuszkiewicz 2019, Feng *et al.* 2012, Melo *et al.* 2004). Interestingly, respiratory Complex I was lost several times during the evolution of Ascomycota yeasts, and as a result is not present in several well-established yeast genera such as *Saccharomyces*, *Kluyveromyces* and *Schizosaccharomyces* (Dujon 2010, Hagman *et al.* 2013, Riley *et al.* 2016). Species from these genera instead completely rely on alternative NADH dehydrogenase for respiratory oxidation of NADH, of which for example *S. cerevisiae* possesses one internal and two external variants (de Vries and Marres 1987, Luttik *et al.* 1998). The Crabtree-negative yeasts *O. polymorpha* and *O. parapolymorpha* encode both respiratory Complex I and multiple alternative NADH dehydrogenases and activity has been

demonstrated for both systems with submitochondrial particles, suggesting these yeasts possess both respiratory systems in parallel (Bridges *et al.* 2009, Riley *et al.* 2016).

Efficiency of respiratory energy coupling. The energetic efficiency of the overall respiratory process can be expressed by the P/O ratio, which indicates the number of ATP molecules formed by the respiratory chain per pair of electrons ($2e^-$, one NADH equivalent) transferred to one oxygen atom. As such, the P/O ratio is affected by both the stoichiometry of proton translocation of the respiratory chain complexes as electrons pass ($H^+/2e^-$), as well as the number of protons required by the ATPase for formation of one ATP (expressed either as H^+/ATP, or the reciprocal ATP/H^+) (Ferguson 2010). As the alternative NADH dehydrogenases do not couple NADH oxidation to proton translocation and therefore exhibit lower $H^+/2e^-$ ratios compared to respiratory Complex I, NADH oxidation via these enzymes results in a lower P/O ratio (Kerscher 2000). Since Complex I is responsible for 4 out of the 10 protons that are translocated by the respiratory chain per 1 NADH oxidized, the use of alternative NADH dehydrogenase entails a reduction in ATP produced per NADH oxidized and, by extension, per molecule of substrate dissimilated and oxygen consumed.

P/O ratios of mitochondrial oxidative phosphorylation have long been a topic of debate (Ferguson 2010). Originally it was thought that each coupling site, i.e. complexes I, III and IV, would result in 1 ATP per pair of electrons passed. This assumption resulted in a P/O ratio of 3 for NADH oxidation via Complex I and a P/O ratio of 2 for NADH oxidation via internal or external alternative NADH dehydrogenase[2] (Silverstein 2005). Under these assumptions, with a respiratory chain that utilizes Complex I for oxidation of mitochondrial NADH and alternative NADH dehydrogenase for oxidation of cytosolic NADH, complete respiratory dissimilation of 1 glucose would yield 32 ATP from oxidative phosphorylation with an overall P/O ratio of 2.67 (32 ATP divided by 12 O atoms consumed) and an additional 4 ATP from SLP for a total of 36 ATP. With the same assumptions but lacking Complex I and using internal alternative NADH dehydrogenase instead for oxidation of mitochondrial NADH, the resulting overall P/O ratio would be 2, yielding a lower total of 28 ATP (24 + 4) per 1 glucose respired (**Table 1.1**).

As biochemical methods, accuracy of measuring P/O ratios and knowledge about respiratory coupling improved, the insight that P/O ratios are affected by the H^+ stoichiometries of proton-pumping respiratory complexes as well as the H^+/ATP ratio of ATPase led researchers to abandon the notion that P/O ratios of coupling sites (i.e. of respiratory complexes) had to be integral numbers. Based on experimental data indicating a H^+/ATP ratio of 3 for ATPase, and taking into account an additional proton

[2] And also, a P/O ratio of 2 for oxidation of succinate via Complex II (succinate dehydrogenase).

per ATP due to the net cost for transport of ADP/ATP/Pi across the IMM (effective H^+/ATP = 4) (Hinkle 2005, Hinkle *et al.* 1991, Silverstein 2005), overall P/O ratios of 2.17 and 1.5 can be calculated for organisms with and without Complex I, respectively[3] (**Table 1.1**).

An additional improvement in calculation of P/O ratios came from structural analysis of mitochondrial ATPases and was the realization that the number of c subunits that comprises the c-ring of ATPase determines the number of protons required for one cycle of ATPase (producing 3 ATP), thereby specifying the H^+/ATP ratio (Ferguson 2010). In ATPases from yeasts, the c-ring has been determined to contain 10 c subunits (Stock *et al.* 1999, Vinothkumar *et al.* 2016), which results in an effective H^+/ATP ratio of 4.33 and further adjusts calculated overall P/O ratios to 2 and 1.4 for species with and without Complex I, respectively (**Table 1.1**). Intriguingly, the ATPase c-ring size is not conserved between different domains of life, and the c-ring size of 8 (effective H^+/ATP = 3.67) determined for ATPase from bovine heart mitochondria (Watt *et al.* 2010), suggests that animal cells have an intrinsic advantage in the efficiency of respiratory energy coupling over yeasts at the level of ATPase.

It is important to note that P/O ratios of microbial respiratory chains, calculated based on structural considerations, cannot (yet) be reconciled with *in vivo* determined P/O ratios (also referred to as operational P/O ratios) that are obtained by quantitative analysis of cellular growth. For yeasts and fungi, determined *in vivo* P/O ratios are typically lower than their respective predicted P/O ratios (**Table 1.1**), as they factor in cellular processes outside the respiratory chain, e.g., the use of the proton gradient for mitochondrial protein import (Gasser *et al.* 1982) and oxygen consumption by reactions other than the respiratory chain (Nes 2011).

[3] These assumptions are the basis for the P/O ratios of 2.5 and 1.5 that are typically reported in textbooks for Complex I and succinate dehydrogenase/alternative NADH dehydrogenase, respectively.

Assumptions			**Resulting P/O ratio at coupling sites**		**Calculated P/O ratio and resulting ATP (in brackets) produced from 1 molecule of respired glucose**	
Coupling	**H^+/ATP (ATPase)**	**H^+/ATP (transport)**	**Complex I**	**NDH2**	**With Complex I**	**Without Complex I**
1 ATP per pass of NADH eq. through respiratory complex	n/a	n/a	3	2	2.67 (32+4)	2 (24+4)
10 and 6 total H^+ translocated for entry of NADH eq. at Complex I and NDH2, respectively	3	1	2.50	1.50	2.17 (26+4)	1.50 (18+4)
10 and 6 total H^+ translocated for entry of NADH eq. at Complex I and NDH2, respectively	3.33	1	2.31	1.38	2.00 (24+4)	1.38 (16.62+4)
n/a			unknown		***In vivo* determined P/O ratio and resulting ATP (in brackets) produced from 1 molecule of respired glucose**	
					With Complex I	**Without Complex I**
					Candida utilis 1.53* (18.36+4) (van Gulik and Heijnen 1995) *Penicillium chrysogenum* 1.84 (22.08+4) (van Gulik *et al.* 2001)	*Saccharomyces cerevisiae* 1.18 (14.16+4) (Bonnet *et al.* 1980) 1.09 (13.08+4) (Vanrolleghem *et al.* 1996) 0.95 (11.40+4) (Verduyn *et al.* 1991)

* Compartmentation not considered – underestimation

Table 1.1: **P/O ratios of aerobic respiration based on assumption-based calculation and *in vivo* determination**. Based on respiratory chain that oxidizes cytosolic NADH via external NADH dehydrogenase and mitochondrial NADH via Complex I (‘ with Complex I’) or internal NADH dehydrogenase (‘ Without Complex I’). Calculations of P/O ratios based on respiration of 8 mitochondrial NADH, 2 mitochondrial $FADH_2$, and 4 cytosolic NADH with P/O ratio of $FADH_2$ oxidation (Complex II) set equal to NADH oxidation by NDH2. Resulting ATP by one molecule of respired glucose calculated by multiplication of P/O ratio with 12 (number of NADH or $2e^-$ equivalent from one glucose) + 4 (ATP from substrate level phosphorylation). Calculated P/O ratios and ATP amounts rounded to two decimal places in case of non-integer numbers. n/a, not applicable.

Genetic engineering of yeasts. Effective and efficient genetic engineering is a crucial tool for studying cellular metabolism and/or altering strain properties towards efficient product formation, and has seen constant development over the last 50 years. Early developments, such as the restriction (Arber and Dussoix 1962, Arber and Linn 1969) and ligation (Weiss and Richardson 1967, Zimmerman *et al.* 1967) of DNA, first enabled the development of recombinant DNA technology for *E. coli* and yielded the first recombinant microorganism (Cohen *et al.* 1973). Afterwards, the development of yeast transformation protocols by use of a plasmid with homology sequences (Hinnen *et al.* 1978) and work on 'recombination' in yeast using the double strand break (DSB) model (Orr-Weaver *et al.* 1981, Szostak *et al.* 1983) led to the modification of *S. cerevisiae* by numerous researchers and the development of various methods and tools for yeast genetic engineering (Fraczek *et al.* 2018). Since then, improvements in yeast genetic engineering included the development of single-step gene disruption by use of selection markers flanked by homologous sequences (Rothstein 1983), the development of PCR (Mullis *et al.* 1992) which removed the need for a cloning step in *E. coli* by using oligonucleotides, resembling chromosomal targets, for *in vitro* amplification of transformation fragments (Baudin *et al.* 1993). Also, standardized sets of cassettes and plasmids for use by a wide range of yeast researchers were developed (Wach *et al.* 1994). The observation that overlapping DNA fragments can interact and integrate at a single locus *in vivo* (Plessis and Dujon 1993) allowed the development of split-marker recombination (Fairhead *et al.* 1996), further reducing the steps needed to achieve gene deletion and integration and allowing combinatorial approaches. The split-marker technique also proved useful for targeted genetic alteration of hosts with low frequency of homologous recombination. By not transforming the marker cassette as a single fragment, occurrences of random, multiple and/or tandem integration can be reduced and the likelihood of correct recombination-based genomic integration increased (**Figure 1.4A**) (Chung and Lee 2015).

Another development was the concept of 'marker recycling', where after successful modifications genetic markers were removed from the mutant genome. This allowed re-use of the same genetic marker, alleviating a bottleneck in genome engineering projects requiring multiple genetic alterations (Fraczek *et al.* 2018, Goffeau *et al.* 1996). Marker recycling was realized by tools such the cre-*lox*P system (Delneri *et al.* 2000, Sauer 1987), the *delitto perfetto* method (Storici *et al.* 2001) and meganuclease-based methods such as the use of I-SceI to induce recombination-stimulating DSBs in the yeast genome (Storici *et al.* 2003).

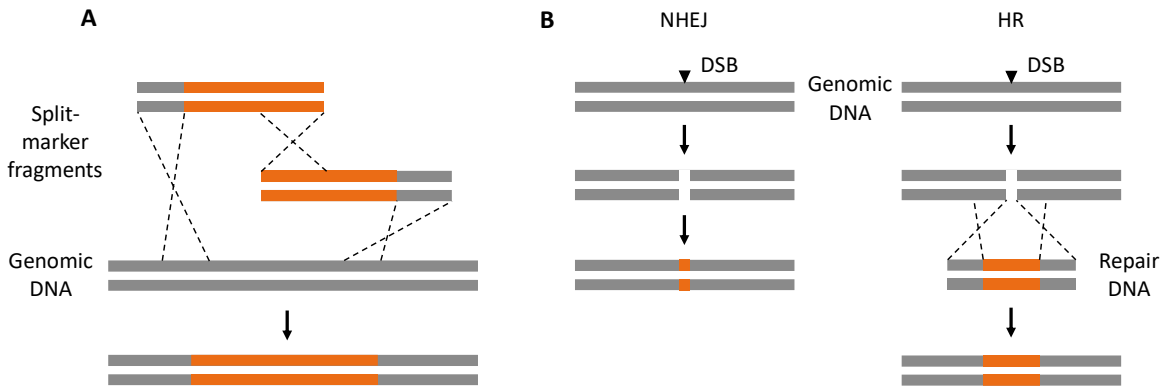


Figure 1.4: Split marker recombination and double strand break (DSB) repair. **A**: for a split marker transformation, the selection marker (orange) is transformed as two split fragments with internal homology. Before integration into the genome, the fragments have to recombine internally. **B**: DSB repair is mediated either by error-prone non-homologous end joining (NHEJ) leading to the integration of small indels in the genome, or by homologous recombination (HR) with an externally supplied repair DNA fragment which allows insertion, replacement or deletion of larger sequences.

In light of these developments, the Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR) and CRISPR-associated protein 9 (Cas9) system have revolutionized yeast genome editing over the past years (Fraczek *et al.* 2018). This revolutionary system works via two components: the Cas9 endonuclease (typically from *Streptococcus pyogenes*) and a guide RNA (gRNA) molecule that binds to Cas9 to form a protein-RNA complex. If this complex encounters a DNA sequence that is complementary to the variable target sequence included in the gRNA and an 'NGG' protospacer adjacent motif (PAM) sequence is present immediately next to the target sequence within the DNA target, Cas9 induces a DSB (Jinek *et al.* 2012). A DSB is lethal unless repaired by cellular DNA mechanisms, which in yeasts is predominantly accomplished by two pathways: homologous recombination (HR), or error-prone non-homologous end joining (NHEJ), depending of on the presence of DNA fragments that could serve as repair template and the host cell's predominant mechanism of DSB repair (**Figure 1.4B**) (Shrivastav *et al.* 2008). While error-prone NHEJ allows the disruption of genes by small insertion or deletion (indel) mutations, DSB repair via HR allows for powerful genome editing by insertion or replacement of sequences at the targeted locus via the use of donor DNA, enabling marker-free and scarless genome editing (Cai *et al.* 2019). After the first description of CRISPR-Cas9 as a genome editing system, a tool adapted to *S. cerevisiae* quickly emerged (DiCarlo et al. 2013) and was subsequently also adopted for use in non-conventional yeast species (Löbs *et al.* 2017b).

Scope of this thesis. This PhD project was conducted at the Delft University of Technology in the framework of the BE-Basic foundation, a public-private partnership aimed at advancing the biobased economy. The objectives of the research described in this thesis were the investigation and engineering of respiratory energy coupling in

yeasts, with the goal of enabling rational design of strains and process conditions for improving aerobic bioprocesses. The popular and widely employed yeast *S. cerevisiae*, due to its lack of a proton-translocating respiratory Complex I NADH dehydrogenase, is inherently limited in its efficiency to couple substrate oxidation to generation of ATP via the respiratory chain. The focus of this project was therefore instead on the non-conventional yeasts *O. polymorpha* and *O. parapolymorpha*, which do possess Complex I, are established production hosts for recombinant proteins and combine a range of other industrially relevant properties such as thermotolerance and methylotrophy.

Efficient and effective genome editing tools are helpful in gaining understanding of cellular mechanisms and the design and construction of yeast cell factories. In recent years, the CRISPR-Cas9 system has transformed genetic editing in yeast, typically with tools that are highly species-specific. **Chapter 2** of this thesis describes the development and experimental evaluation of a multispecies CRISPR-Cas9 tool for application in *Ogataea* and *Kluyveromyces* yeasts.

Quantitative physiology research requires the establishment of a baseline of strain physiology and performance. **Chapter 3** describes a comprehensive physiological characterization of two industrially-relevant strains of *O. polymorpha* and *O. parapolymorpha* under defined, fully aerobic conditions in bioreactor-grown batch- and chemostat cultures. As this type of work requires complex, but very standardized methodology, it was combined with the evaluation of a software package for definition of experimental procedures and linked data analysis.

The data gathered on aerobic physiology of *O. polymorpha* and *O. parapolymorpha* indicated that, despite initial assumptions, respiratory Complex I is not physiologically relevant in these yeasts under the investigated cultivation conditions. **Chapter 4** therefore aims at the quantitative dissection of the contribution of Complex I to respiratory metabolism in *O. parapolymorpha*. This objective was approached by use of a Complex I-deficient mutant strain, systems biology methods to track expression patterns, and quantitative physiology experiments at a variety of specific growth rates, including retentostat cultivation which allowed investigation under extreme substrate limitation at near-zero specific growth rates.

The investigation of Complex I revealed that under some conditions, alternative NADH dehydrogenases are used for NADH oxidation in *O. parapolymorpha*, which has implications for yields in aerobic bioprocesses and offers opportunities for metabolic engineering. In **Chapter 5**, the physiological relevance, mitochondrial localization and substrate specificity of these enzymes in *O. parapolymorpha* were investigated by a combination of quantitative physiology experiments with growing cultures and isolated mitochondria, as well as by *in vitro* characterization after heterologous expression in *Escherichia coli*.

Chapter 2

Genome editing in *Kluyveromyces* and *Ogataea* yeasts using a broad-host-range Cas9/gRNA co-expression plasmid

Hannes Jürgens
Javier A. Varela
Arthur R. Gorter de Vries
Thomas Perli
Veronica J.M. Gast
Nikola Y. Gyurchev
Arun S. Rajkumar
Robert Mans
Jack T. Pronk
John P. Morrissey
Jean-Marc G. Daran

Essentially as published in FEMS Yeast Research (2018); 18(3):foy012
https://doi.org/10.1093/femsyr/foy012

Abstract

While CRISPR-Cas9-mediated genome editing has transformed yeast research, current plasmids and cassettes for Cas9 and guide-RNA expression are species specific. CRISPR tools that function in multiple yeast species could contribute to the intensifying research on non-conventional yeasts. A plasmid carrying a pangenomic origin of replication and two constitutive expression cassettes for Cas9 and ribozyme-flanked gRNAs was constructed. Its functionality was tested by analyzing inactivation of the *ADE2* gene in four yeast species. In two *Kluyveromyces* species, near-perfect targeting (≥96%) and homologous repair (HR) were observed in at least 24% of transformants. In two *Ogataea* species, Ade$^-$ mutants were not observed directly after transformation, but prolonged incubation of transformed cells resulted in targeting efficiencies of 9% to 63% mediated by non-homologous end joining (NHEJ). In an *Ogataea parapolymorpha ku80* mutant, deletion of *OpADE2* mediated by HR was achieved, albeit at low efficiencies (<1%). Furthermore, the expression of a dual polycistronic gRNA array enabled simultaneous interruption of *OpADE2* and *OpYNR1* demonstrating flexibility of ribozyme-flanked gRNA design for multiplexing. While prevalence of NHEJ prevented HR-mediated editing in *Ogataea*, such targeted editing was possible in *Kluyveromyces*. This broad-host-range CRISPR/gRNA system may contribute to exploration of Cas9-mediated genome editing in other *Saccharomycotina* yeasts.

Introduction

Design and construction of efficient yeast cell factories for industrial production of fuels and a wide range of chemicals is among the key developments in microbial biotechnology in the last 20 years. *Saccharomyces cerevisiae* is currently the most intensively used yeast species for production of low-molecular-weight products such as alcohols, organic acids and isoprenoid. Its tolerance to high sugar concentrations and low pH, its overall robustness, its ability to grow anaerobically and especially its tractability to genetic manipulation make *S. cerevisiae* a convenient chassis for various biotechnological purposes (Kavscek *et al.* 2015, Li and Borodina 2015). Furthermore, availability of high-quality genome sequences and well-developed genetic tools (Salazar *et al.* 2017, Stovicek *et al.* 2017) facilitate tailoring of *S. cerevisiae* to specific industrial applications, either by improving existing traits or by expressing heterologous enzymes and pathways. However, its narrow temperature spectrum, limited substrate range and strong tendency toward alcoholic fermentation under aerobic conditions have stimulated interest in studying non-*Saccharomyces* species, often referred to as 'non-conventional yeasts', that exhibit attractive features for further improving sustainability and economic viability of biobased production (Johnson 2013).

Kluyveromyces lactis and *K. marxianus* are non-conventional yeasts that are generally regarded as safe (GRAS), can utilize lactose as a carbon source and have been applied for bioethanol production from cheese whey, a by-product of the dairy industry (Siso 1996). *Kluyveromyces lactis* has an excellent capacity for protein secretion, which has been exploited for production of several heterologous proteins (Spohner *et al.* 2016). *Kluyveromyces marxianus* can grow very fast at temperatures above 40°C and has been used for the production of bioethanol, biomass and flavor compounds (Morrissey *et al.* 2015). The thermotolerant and methylotrophic yeast *Ogataea polymorpha* (syn. *Hansenula polymorpha*) is a major established platform for heterologous protein expression thanks to the availability of extremely strong yet tightly controlled promoters (Gellissen 2000). In addition, this yeast has been engineered for high-temperature ethanol fermentation from various carbon sources, including xylose (Kurylenko *et al.* 2014), and glycerol, a by-product originating from biodiesel production (Kata *et al.* 2016). *Ogataea polymorpha* shares many characteristics with the closely related species *O. parapolymorpha*, which was taxonomically separated from *O. polymorpha* in 2010, and now includes the popular former *H. polymorpha* DL-1 strain (Suh and Zhou 2010).

The unique physiological characteristics of non-conventional yeast species have the potential to reduce production costs of processes that are currently performed with *S. cerevisiae*. However, analogous to the situation in industrial *S. cerevisiae* strains, this will require extensive genetic engineering. While obtaining the necessary genome-

sequence information is relatively straightforward due to fast developments in whole-genome sequencing (Goodwin *et al.* 2016), genetic modification by classical methods is still challenging mostly due to the predominant DNA-repair mechanisms of many non-conventional yeasts (Abdel-Banat *et al.* 2010, Nonklang *et al.* 2008). Unlike *S. cerevisiae*, these yeasts typically have a very active non-homologous end joining (NHEJ) DNA repair mechanism and use homologous recombination (HR) to a much lesser extent, which makes precise genome editing inefficient as provided linear repair DNA fragments are not efficiently integrated at the targeted genomic locus (Klinner and Schäfer 2004). Introduction of a double-strand break (DSB) in the targeted DNA locus can strongly facilitate genetic engineering, either by introduction of mutations at the cut site or by stimulating the occurrence of HR-mediated DNA repair with co-transformed repair fragments (Jasin and Rothstein 2013, Kuijpers *et al.* 2013a).

Over the past five years, the CRISPR-Cas9 system has emerged as a powerful and versatile tool to engineer the genomes of a wide range of organisms (Hsu *et al.* 2014). In this system, the endonuclease CRISPR associated protein 9 (Cas9) binds a guide RNA molecule (gRNA) that targets a sequence-specific site in the genome (Jinek *et al.* 2012). The Cas9-gRNA complex then induces a DSB that is lethal unless repaired. Repair of DSBs typically occurs through NHEJ or HR, depending on the presence of a repair DNA fragment and the predominant DSB repair mechanism of the host cell (Shrivastav *et al.* 2008). CRISPR-Cas9's ability to induce mutations in the target sequence has been widely exploited in the development of genetic tools for various non-conventional yeasts including *K. lactis* (Horwitz *et al.* 2015), *K. marxianus* (Löbs *et al.* 2017a, Nambu-Nishida *et al.* 2017), *O. polymorpha* (Numamoto *et al.* 2017), *Pichia pastoris* (Weninger *et al.* 2016), *Scheffersomyces stipitis* (Cao *et al.* 2017) and *Yarrowia lipolytica* (Gao *et al.* 2016, Schwartz *et al.* 2016). However, the plasmid and Cas9/gRNA expression systems available today have generally only been designed for and tested in a single yeast species, thus limiting the potential of each system to work on different yeasts with novel characteristics.

In this study, we developed a novel wide-host-range CRISPR-Cas9 system for use across several yeast species, based on plasmid-borne expression of Cas9 and a ribozyme-flanked gRNA that was recently developed for *S. pastorianus*, a *Saccharomyces* hybrid highly resilient to genetic modification (Gorter de Vries *et al.* 2017). The system was successfully tested in *K. marxianus*, *K. lactis*, *O. polymorpha* and *O. parapolymorpha*. The results highlight the potential of cross-species CRISPR-Cas9 tools for genome engineering in yeasts.

Materials and Methods

Strains and growth conditions. The *K. lactis*, *K. marxianus*, *O. polymorpha* and *O. parapolymorpha* strains used in this study are listed in **Table 2.1**. For cultivation under non-selective conditions, strains were grown in 500 mL shake flasks containing 100 mL YPD medium (10 g L^{-1} Bacto yeast extract, 20 g L^{-1} Bacto peptone, 20 g L^{-1} glucose, demineralized water), placed in a rotary shaker set to 30°C and 200 rpm. For antibiotic selection, media for cultivation of *Kluyveromyces* and *Ogataea* species were supplemented with 200 and 300 µg mL^{-1} hygromycin B, respectively. Prolonged liquid incubation for generation of gene disruptions in *Ogataea* species was carried out under selective conditions in 50 mL vented tubes (Cellstar CELLreactor, Greiner Bio-One, Kremsmünster, Austria), containing 25 mL YPD medium and incubated as described above. Solid medium was prepared by addition of 2% (w/v) agar. To verify the disruption of nitrate reductase gene *YNR1*, *O. parapolymorpha* was grown on synthetic medium (SM) which contained 20 g L^{-1} glucose, 3 g L^{-1} KH_2PO_4, 0.5 g L^{-1} $MgSO_4 \cdot 7\ H_2O$, 5 g L^{-1} $(NH_4)_2SO_4$, 1 mL L^{-1} of a trace element solution and of a vitamin solution as previously described in Verduyn *et al.* (1992) and on synthetic medium with nitrate (SMN) in which $(NH_4)_2SO_4$ was substituted with 5 g L^{-1} K_2SO_4 and 4.3 g L^{-1} $NaNO_3$. For multiplexed targeting of *ADE2* and *YNR1* in *O. parapolymorpha*, SM and SMN media were supplemented with 15 mg L^{-1} adenine to allow for growth of Ade^- mutants. Frozen stock cultures were prepared from exponentially growing shake-flask cultures by addition of 30% (v/v) glycerol, and aseptically stored in 1 mL aliquots at -80°C.

Construction of plasmids and repair DNA fragments. All plasmids used in this study are described in **Table 2.2**. The DNA parts harbored by plasmids pUD527, pUD530, pUD531, pUD532, pUD540, pUD555, pUD602 and pUD750 were *de novo* synthesized by GeneArt (Thermo Fisher Scientific, Waltham, MA, USA).

For the construction of pUDP002, an intermediate plasmid containing the *Klebsiella pneumoniae hph* (HygR) open reading frame (ORF), expressed from the *TEF1* promoter and terminator from the yeast *Ashbya gossypii* (*Eremothecium gossypii*), was constructed by 'Gibson' assembly (Gibson *et al.* 2009) from plasmids pYTK079 (Lee *et al.* 2015) and pUD527: the *hph* ORF was amplified from pYTK079 using primers 9837 and 9838, and inserted into a plasmid backbone which was generated by PCR amplification of pUD527 using primers 9839 and 9840.

Table 2.1: *Kluyveromyces* and *Ogataea* strains used in this study.

Species	Strain	Genotype	Origin
Kluyveromyces lactis	CBS 2359	Wild type	CBS-KNAW[A], van der Walt (1971)
	IMK829	*Δade2*	This study
Kluyveromyces marxianus	CBS 5795	Wild type	CBS-KNAW[A], van der Walt (1971)
	NBRC 1777	Wild type	NBRCB
	IMK830	*Δade2*	This study
	CBS 397	Wild type	CBS-KNAW[A], van der Walt (1971)
Ogataea polymorpha (syn. *Hansenula polymorpha*)	CBS 4732	Wild type	CBS-KNAW[A], Yamada *et al.* (1994)
Ogataea parapolymorpha (syn. *Hansenula polymorpha*)	CBS 11895 (DL-1, ATCC 26012)	Wild type	CBS-KNAW[A], Suh and Zhou (2010)
	IMD001	*ku80*^{A340AA}	This study
	IMK828	*ku80*^{A340AA} *Δade2*	This study
	IMD034	*ade2*^{C120CA} *ynr1*^{G397GT}	This study

CBS 2359, CBS 5795, CBS 397, CBS 4732 and CBS 11895 were obtained from the CBS-KNAW fungal collection ([A] Westerdijk Fungal Biodiversity Institute, Utrecht, The Netherlands). NBRC 1777 was obtained from the NBRC culture collection ([B] National Institute of Technology and Evaluation, Tokyo, Japan).

Plasmid pUDP002 was constructed by Gibson assembly from five overlapping fragments, using synthetic homologous recombination sequences (Kuijpers *et al.* 2013b): (i) the *AgTEF1p-hph-AgTEF1t* hygromycin resistance cassette was amplified from the intermediate plasmid using primers 9841 and 9842 and sequence-verified before further use; (ii) the *AaTEF1p-cas9*$^{D147Y\ P411T}$*-ScPHO5t* expression cassette (Bao *et al.* 2015, Gorter de Vries *et al.* 2017) was obtained by amplification of pUD423 using primers 3841 and 9393; (iii) the pangenomic yeast replication origin panARS(OPT) (Liachko and Dunham 2014) was amplified from pUD530 using primers 9663 and 3856; (iv) the empty *ScTDH3p-ScCYC1t* gRNA expression cassette was obtained by digestion of pUD531 with SmaI; (v) the *Escherichia coli* pBR322 origin and *bla* gene for ampicillin selection were obtained by digestion of pUD532 with SmaI. The five fragments were gel-purified and 0.1 pmol of each fragment was assembled in a Gibson assembly reaction. An *E. coli* clone containing the correctly assembled plasmid (verified by digestion with PdmI) was stocked and pUDP002 (Addgene Plasmid #103872) was used as gRNA entry plasmid for subsequent plasmid construction.

The gRNA expression plasmids pUDP013 (plasmid #103873), pUDP025 (plasmid #103874), pUDP046 (plasmid #107062) and pUDP123 (plasmid #107269) were assembled in a one-pot 'Golden Gate' assembly (Engler *et al.* 2008) by BsaI digestion of pUDP002 and a synthesized donor plasmid containing the respective gRNA sequences flanked by ribozymes and BsaI restriction sites.

Table 2.2: Plasmids used in this study.

Name (Addgene Plasmid #)	Relevant characteristics	Origin
pYTK079	Template for *hph* (Hyg^R) open reading frame	Lee *et al.* (2015)
pUD423	Template for *AaTEF1p-Spcas9*$^{D147Y\ P411T}$*-ScPHO5t* cassette	Gorter de Vries *et al.* (2017)
pUD527	*ori kan*^{R} SHRA *AgTEF1p-amdS-AgTEF1t* SHRB	GeneArt
pUD530	*ori kan*^{R} panARS(OPT)	GeneArt
pUD531	*ori kan*^{R} SHRC *ScTDH3p*-$^{BsaI\ BsaI}$*-ScCYC1t* SHRI	GeneArt
pUD532	*ori kan*^{R} SHRI *ori bla* SHRA	GeneArt
pUD540	*ori amp*$^{R BsaI}$HH-gRNA$_{OpADE2}$-HDV^{BsaI}	GeneArt
pUD555	*ori amp*$^{R BsaI}$HH-gRNA$_{KlADE2}$-HDV^{BsaI}	GeneArt
pUD602	*ori amp*$^{R BsaI}$HH-gRNA$_{OpKU80}$-HDV^{BsaI}	GeneArt
pUD750	*ori amp*$^{R BsaI}$HH-gRNA$_{OpADE2}$-HDV-HH-gRNA$_{OpYNR1}$-HDV^{BsaI}	This study
pUDP002 (plasmid #103872)	*ori amp*^{R} panARS(OPT) *AgTEF1p-hph-AgTEF1t ScTDH3p*$^{BsaI\ BsaI}$*ScCYC1t AaTEF1p-Spcas9*$^{D147Y\ P411T}$*-ScPHO5t*	This study
pUDP013 (plasmid #103873)	*ori amp*^{R} panARS(OPT) *AgTEF1p-hph-AgTEF1t ScTDH3p*-HH-gRNA$_{OpADE2}$-HDV- *ScCYC1t AaTEF1p-Spcas9*$^{D147Y\ P411T}$*-ScPHO5t*	This study
pUDP025 (plasmid #103874)	*ori amp*^{R} panARS(OPT) *AgTEF1p-hph-AgTEF1t ScTDH3p*-HH-gRNA$_{KlADE2}$-HDV- *ScCYC1t AaTEF1p-Spcas9*$^{D147Y\ P411T}$*-ScPHO5t*	This study
pUDP046 (plasmid #107062)	*ori amp*^{R} panARS(OPT) *AgTEF1p-hph-AgTEF1t ScTDH3p*-HH-gRNA$_{OpKU80}$-HDV- *ScCYC1t AaTEF1p-Spcas9*$^{D147Y\ P411T}$*-ScPHO5t*	This study
pUDP082 (plasmid #103875)	*ori amp*^{R} panARS(OPT) *AgTEF1p-hph-AgTEF1t ScTDH3p*-HH-gRNA$_{KmADE2}$-HDV- *ScCYC1t AaTEF1p-Spcas9*$^{D147Y\ P411T}$*-ScPHO5t*	This study
pUDP123 (plasmid # 107269)	*ori amp*^{R} panARS(OPT) *AgTEF1p-hph-AgTEF1t ScTDH3p*-HH-gRNA$_{OpADE2}$-HDV-HH-gRNA$_{OpYNR1}$-HDV-*ScCYC1t AaTEF1p-Spcas9*$^{D147Y\ P411T}$*-ScPHO5t*	This study

Restriction enzyme sites are indicated in superscript and gRNA target sequences are indicated in subscript. SHRs represent specify synthetic homologous recombination sequences used for plasmid assembly. Aa: *Arxula adeninivorans;* Sp: *Streptococcus pyogenes;* Ag: *Ashbya gossypii*; Sc: *Saccharomyces cerevisiae*; Op: *Ogataea* (*para*)*polymorpha*; Kl: *Kluyveromyces lactis*; Km: *Kluyveromyces marxianus*; HH: hammerhead ribozyme; HDV: hepatitis delta virus ribozyme. pUDP013 (gRNA$_{OpADE2}$) targets *ADE2* in both *O. polymorpha* and *O. parapolymorpha*. The Addgene plasmid code (when relevant) is indicated next to the plasmid name between brackets.

pUDP013, pUDP025, pUDP046 and pUDP123 were constructed using pUD540, pUD555, pUD602 and pUD750, respectively. The Golden Gate assemblies were carried out in a final volume of 20 µL, using T4 DNA ligase buffer (Thermo Fisher Scientific), 10 U BsaI (New England Biolabs), 1 U T4 DNA ligase (Thermo Fisher Scientific) and 10 ng of both pUDP002 and the respective donor plasmid. For the construction of pUDP082 (plasmid #103875), the partially overlapping primers Km-ade2-F and Km-ade2-R were annealed, extended and amplified by PCR to yield a 233 bp double-stranded DNA fragment containing the *KmADE2* gRNA target sequence flanked by ribozymes and BsaI sites, which was integrated into pUDP002 by Golden Gate

assembly as described above. All constructed gRNA-harboring plasmids were verified by digestion with PdmI. Plasmids pUDP002, pUDP013, pUDP025, pUDP046, pUDP082 and pUDP123 were deposited at Addgene (https://www.addgene.org/).

The *ADE2* repair DNA fragments of *K. lactis* (962 bp), *K. marxianus* (959 bp), *O. polymorpha* (960 bp) and *O. parapolymorpha* (960 bp) were constructed from strains CBS 2359, CBS 5795, CBS 4732 and CBS 11895, respectively, using primer sets 10723 & 10724 and 10725 & 10726 (*K. lactis*), 10727 & 10728 and 10729 & 10730 (*K. marxianus*), 10346 & 10347 and 10348 & 10349 (*O. polymorpha*) and 10354 & 10355 and 10356 & 10357 (*O. parapolymorpha*) to amplify the homology regions flanking *ADE2*. Both regions were then joined by overlap extension PCR using primer sets 10723 & 10726, 10727 & 10730, 10346 & 10349 and 10354 & 10357 in the case of *K. lactis*, *K. marxianus*, *O. polymorpha* and *O. parapolymorpha*, respectively. In all cases, the final repair fragment was gel-purified and further amplified to obtain quantities required for transformation, using primer sets 10723 & 10726, 10727 & 10730, 10346 & 10349, and 10354 & 10357 for *K. lactis*, *K. marxianus*, *O. polymorpha* and *O. parapolymorpha*, respectively. The primer pairs were designed to amplify the 480 bp (±1 or 2 bp in some cases) upstream the ATG or the 480 bp terminator region downstream the stop codon of the interrupted/deleted ORF.

Design of gRNA target sequences and BsaI-flanked entry constructs. All 20-bp Cas9 gRNA target sequences used in this study are described in **Table 2.3**. Approximately ten candidate target sequences were chosen from the first third of each targeted ORF, based on the presence of an NGG protospacer adjacent motif (PAM) site. Any target sequence with off-targets (defined as a sequence with NGG or NAG PAM and more than 15 nucleotides identical to the candidate sequence) was excluded, based on a blastn homology search against the respective yeast genome (https://blast.ncbi.nlm.nih.gov). The remaining target sequences were ranked based on AT content ('AT score'; 0 being the lowest and 1 being the highest possible AT content) and secondary structure as predicted with the complete gRNA sequence, using minimum free energy prediction by RNAfold (Lorenz *et al.* 2011) ('RNA score'; 0 being the lowest and 1 being the highest possible number of unpaired target sequence nucleotides). Finally, the target sequences with the highest cumulative AT and RNA score that did not exceed an AT score of 0.8 were chosen for use in this study.

To integrate ribozyme-flanked gRNAs into pUDP002, the synthetic 233-bp dsDNA gRNA entry constructs were flanked by inward-facing BsaI sites generating sticky ends (underlined) 'GGTCTCGCAAA' and 'ACAGCGAGACC' at their 5' and 3' ends, respectively, compatible with BsaI-digested pUDP002. The sequence between the BsaI sites consisted of (i) the hammerhead ribozyme with the first six nucleotides being the reverse complement (^{c}) of the first six nucleotides of the gRNA spacer

'$^{c}N_6{}^{c}N_5{}^{c}N_4{}^{c}N_3{}^{c}N_2{}^{c}N_1$CUGAUGAGUCCGUGAGGACGAAACGAGUAAGCUCGUC', (ii) followed by the 20-nucleotide gRNA spacer followed by the structural gRNA 'GUUUUAGAGCUAGAAAUAGCAAGUUAAAAUAAGGCUAGUCCGUUAUCAACUUGAAAAAG UGGCACCGAGUCGGUGCUUUU', (iii) followed by the hepatitis delta virus ribozyme 'GGCCGGCAUGGUCCCAGCCUCCUCGCUGGCGCCGGCUGGGCAACAUGCUUCGGCAUGGCGA AUGGGAC' (Gao and Zhao 2014, Gorter de Vries *et al.* 2017). For multiplexed targeting of *O. parapolymorpha ADE2* and *YNR1* using a single expression cassette, two ribozyme-flanked gRNAs were connected in a tandem array using 20-bp linker 'GTGTAATGTCCAGAGTTGTG', and otherwise constructed as described previously (Gorter de Vries *et al.* 2017).

Table 2.3: gRNA target sequences used in this study.

Locus	Target sequence (5′-3′)	Position in ORF (bp)	AT score	RNA score
KlADE2	TTTCAATACCTCAAGTGCTTCGG	508/1710	0.65	0.70
KmADE2	GCCCATTTTTCTGCGTATAGCGG	537/1710	0.55	0.70
*OpADE2**	CTGGAATTGATCTGCTTGGCCGG	120/1704	0.50	0.35
OpKU80	CATCGTTCTGCAGAAGATCATGG	340/2076	0.55	0.55
OpYNR1	AGCACAGACCATAGTAACTGGGG	397/2580	0.55	0.55

Target sequences are shown including PAM sequence (underlined). The gRNA for *OpADE2* targets the respective genes at the same position in both *O. polymorpha* and *O. parapolymorpha*. Position in ORF indicates the base pair after which the Cas9-mediated DSB is expected, out of the total length of the ORF. AT score indicates AT content of the 20-bp target sequence. RNA score indicates the fraction of unpaired nucleotides of the 20-bp target sequence, predicted with the complete gRNA sequence using minimum free energy prediction by *RNAfold* (Lorenz *et al.*2011). *The same sequence was used for single (pUDP013) and for double (pUDP123) editing.

Strain construction. Yeast transformations: *Kluyveromyces lactis* and *K. marxianus* were transformed using the LiAc/single-strand carrier DNA/polyethylene glycol method (Gietz and Schiestl 2007). Overnight pre-cultures in YPD medium were used to inoculate a shake flask containing YPD medium to an initial OD_{660nm} of 0.5. Cultures were then incubated at 30°C until an OD_{660nm} of 2.0 was reached, harvested and transformed with 200 ng of plasmid DNA and 300 or 1000 ng of repair DNA fragment in the case of *K. lactis* or *K. marxianus*, respectively. After heat shock, cells were recovered in 1 mL YPD medium for 3 h and plated on selective YPD medium. Plates were typically kept for 5 d at 30°C and then incubated at 4°C for 2 h before assessing the percentage of red Ade^- colonies. *Ogataea polymorpha* and *O. parapolymorpha* were transformed using the procedure for preparation of competent cells and electroporation described by Saraya *et al.* (2014), with modifications. All steps were performed at 30°C, OD_{660nm} of all cultures harvested for transformation was normalized to OD_{660nm} 1.2 by dilution with sterile demineralized water, all centrifugation steps were carried out for 3 min at 3000 g, the DTT incubation step was done for 20 min, and the washing step with STM buffer was performed twice with 50 mL. Electroporation was carried out with 40 µL of a freshly prepared competent cell suspension in 2 mm

gap cuvettes (Bio-Rad, Hercules, CA, USA), using 200 ng of plasmid DNA and 1 µg of repair fragment. A Micropulser Electroporator (Bio-Rad) set to the 'Sc2' preset (1.5 kV) was used for pulse delivery. After electroporation, cells were recovered in 1 mL YPD medium for 1 h at 30°C before plating onto selective YPD medium. For the direct inoculation of prolonged liquid incubation cultures, 100 µL of recovered cell suspension was used as inoculum. Selection plates were typically kept for 4 d at 30°C and then incubated at 4°C for a minimum of 24 h before assessing the percentage of red Ade$^-$ colonies.

Molecular diagnosis of yeast mutants: for molecular analysis of *K. lactis* and *K. marxianus* mutants, colonies were grown overnight in YPD medium. Genomic DNA was extracted using the method described by Singh and Weil (2002), and used as template for PCR reactions targeting the *ADE2* locus. Primer sets 10909 & 10910, and 10911 & 10912 were used for *K. lactis* and *K. marxianus*, respectively. For Sanger sequencing of putative NHEJ-corrected mutants, PCR products were purified and then sequenced using primers 10737 and SeqADE2 for *K. lactis* and *K. marxianus*, respectively. For *O. polymorpha* and *O. parapolymorpha*, genomic DNA was directly isolated from colonies, using the LiAc-sodium dodecyl sulfate method (Lõoke *et al.* 2011). Primer sets 10380 & 10381, and 10915 & 10916 were used to check for deleted *ADE2* loci in *O. polymorpha* and *O. parapolymorpha*, respectively, while sequencing of NHEJ-corrected mutants was done using primer sets 10378 & 10379, and 10386 & 10387, respectively. The deleted *ADE2* locus in strain IMK828 was amplified and Sanger-sequenced using primers 10915 and 10916. Additionally for *O. parapolymorpha* primer pair 12257 & 12266 was used to amplify *OpYNR1* to verify editing of this locus by Sanger sequencing.

Construction of IMK829, IMK830, IMD001, IMK828 and IMD034: to construct IMK829 (*K. lactis ade2*), strain CBS 2359 was co-transformed with 200 ng of pUDP025 (harboring gRNA$_{KlADE2}$) and 300 ng of a 962-bp repair DNA fragment as described above. The resulting transformants were analyzed by diagnostic PCR using primers 10909 and 10910, and a mutant exhibiting a PCR product compatible with deletion of the *ADE2* ORF was isolated and renamed IMK829. To construct IMK830 (*K. marxianus ade2*), strain NBRC 1777 was co-transformed with 200 ng of pUDP082 (harboring gRNA$_{KmADE2}$) and 1 µg of a 959-bp repair DNA fragment as described above. The resulting transformants were analyzed by diagnostic PCR using primers 10911 and 10912, and a mutant exhibiting a PCR product compatible with deletion of the *ADE2* ORF was isolated and renamed IMK830. For the construction of IMD001 (*O. parapolymorpha ku80*^{A340AA}), the *O. parapolymorpha KU80* ORF (accession XM_014078010.1) was identified by a tblastn homology search (https://blast.ncbi.nlm.nih.gov) using the *S. cerevisiae YKU80* protein sequence and *O. parapolymorpha* CBS 11895 (DL-1) RefSeq assembly (accession GCF_000187245.1)

(Ravin *et al.* 2013). Strain CBS 11895 was transformed with 200 ng of pUDP046 (harboring gRNA$_{OpKU80}$), and a single transformant was picked and used to inoculate a prolonged liquid incubation culture as described above. After 96 h, cells were plated on selective YPD medium, genomic DNA was isolated from randomly picked colonies, and the *KU80* locus was amplified and Sanger-sequenced using primers 10751 and 10752. A mutant containing a single adenine nucleotide inserted between position 340 and 341 of the *KU80* ORF was restreaked three times subsequently on non-selective YPD medium to remove pUDP046, and renamed IMD001. To construct IMK828 (*O. parapolymorpha ku80*^{A340AA} *Δade2*), strain IMD001 was co-transformed with 200 ng of pUDP013 (harboring gRNA$_{OpADE2}$) and 1 µg of a 960-bp repair DNA fragment as described above. After recovery, 100 µL of transformation cell suspension was directly used for inoculation of a prolonged liquid incubation culture. After 48 h, cells were plated on selective medium and the resulting colonies inspected for occurrence of the red Ade$^-$ phenotype. A mutant which exhibited a PCR product compatible with deletion of the *ADE2* ORF when analyzed by diagnostic PCR using primers 10915 and 10916 was restreaked thrice subsequently on non-selective YPD medium to remove pUDP013, and renamed IMK828. For the construction of IMD034 (*O. parapolymorpha ade2*^{C120CA} *ynr1*^{G397GT}), the *O. parapolymorpha* HPODL_02384/*YNR1* ORF (accession XM_014082012.1) was identified in *O. parapolymorpha* CBS 11895 (DL-1) RefSeq assembly (accession GCF_000187245.1) as previously described for *OpKU80*. Strain CBS 11895 was transformed with 200 ng of pUDP123 (harboring dual gRNA$_{OpADE2\text{-}OpYNR1}$), and a single transformant was picked and used to inoculate a prolonged liquid incubation culture as described above. After 96 h, cells were plated on selective YPD medium. A set of 94 Ade$^-$ mutants were picked and grown overnight on non-selective YPD and then replica-plated on SM and SMN media. A set of five transformants exhibiting both Ade$^-$ (red colonies) and Nit$^-$ (absence of growth on nitrate) phenotypes were randomly picked, genomic DNA was isolated and the *OpADE2* and *OpYNR1* loci were amplified and Sanger-sequenced using primer pairs 10386 & 10387 and 12257 & 12266 respectively. One sequence-confirmed double interruption mutant was restreaked three times subsequently on non-selective YPD medium to remove pUDP123, and renamed IMD034.

Molecular biology techniques. PCR amplification with Phusion High Fidelity Polymerase (Thermo Fisher Scientific) was performed according to the manufacturer's instructions using PAGE-purified oligonucleotide primers (Sigma-Aldrich, St. Louis, MO, USA). Diagnostic PCR was done using DreamTaq polymerase (Thermo Fisher Scientific) and desalted primers (Sigma-Aldrich). All primer sequences are shown in **Table 2.4**.

Table 2.4: Primers used in this study.

Name	Sequence (5′ - 3′)	Purpose
3841	CACCTTTCGAGAGGACGATG	Construction of pUDP002
3856	CTAGCGTGTCCTCGCATAGTTC	Construction of pUDP002
9393	TGCCGAACTTTCCCTGTATGAAGCGATCTGACCAATCCTTTGCCGTA GTTTCAACGTATGTTTTCATTTTTGCGATGCCAG	Construction of pUDP002
9663	CATACGTTGAAACTACGGCAAAGGATTGGTCAGATCGCTTCATACA GGGAAAGTTCGGCATCAACATCTTTGGATAATATCAGAATGAG	Construction of pUDP002
9837	ATACAGTTCTCACATCACATCCGAACATAAACAAGGATCCATG GGTAAAAAGCCTGAACTC	Construction of pUDP002
9838	ACAAGTTCTTGAAAACAAGAATCTTTTTATTGTCCTCGAGTTATT CCTTTGCCCTCGGAC	Construction of pUDP002
9839	CTCGAGGACAATAAAAAGATTCTTG	Construction of pUDP002
9840	GGATCCTTGTTTATGTTCGGATG	Construction of pUDP002
9841	ACTATATGTGAAGGCATGGC	Construction of pUDP002
9842	GTTGAACATTCTTAGGCTGG	Construction of pUDP002
Km-ade2-F	GGTCTCGCAAAGTCAAGCTGATGAGTCCGTGAGGACGAAACGAGT AAGCTCGTCGCCCATTTTTCTGCGTATAGGTTTTAGAGCTAGAAA TAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTG GCACC	Construction of pUDP082
Km-ade2-R	GGTCTCGCTGTGTCCCATTCGCCATGCCGAAGCATGTTGCCCAGCC GGCGCCAGCGAGGAGGCTGGGACCATGCCGGCCAAAAGCACCGAC TCGGTGCCACTTTTTCAAGTTGATAACGGACTAGCCTTATTTTAA CTTGC	Construction of pUDP082
10723	GTAGTACCGACCTTATCCGTG	Construction of *K. lactis ADE2* repair fragment
10724	GTTGTCTTAGTGAAGAAGGTGAAC	Construction of *K. lactis ADE2* repair fragment
10725	TATATAATAACATCACGTTCACCTTCTTCACTAAGACAACAGCTGC CAAATTAGAAACTATCG	Construction of *K. lactis ADE2* repair fragment
10726	TGTGCGTTGATATATGCCAAC	Construction of *K. lactis ADE2* repair fragment
10727	ATCATAGACAGTCAGTTAGTTCCC	Construction of *K. marxianus ADE2* repair fragment
10728	TTCTTTGGTCCATGATTAACAAGG	Construction of *K. marxianus ADE2* repair fragment
10729	ACTACAACAATATAAACCTTGTTAATCATGGACCAAAGAAGTATTC AACTACCTCCAACAAGAAG	Construction of *K. marxianus ADE2* repair fragment
10730	CAAATTTATGAAGTTTGTGCCATTTG	Construction of *K. marxianus ADE2* repair fragment
10346	TGTGCACTCAATTGCAACC	Construction of *O. polymorpha ADE2* repair fragment
10347	TTCCAACGACCTTTGAGTCC	Construction of *O. polymorpha ADE2* repair fragment
10348	TAATTTAATTTAATTTACATGGACTCAAAGGTCGTTGGAAGTTGGC TATGAGGAATACCTTAAC	Construction of *O. polymorpha ADE2* repair fragment
10349	GGGACGTTTACTGGACGG	Construction of *O. polymorpha ADE2* repair fragment
10354	CCTGATGTGCACTCAATTGC	Construction of *O. parapolymorpha ADE2* repair fragment
10355	CAACGACCTTCGAGTCCATC	Construction of *O. parapolymorpha ADE2* repair fragment
10356	TATTAATTTAATTTAATTTAGATGGACTCGAAGGTCGTTGCTCTG TTGGCTATGAAGAATACC	Construction of *O. parapolymorpha ADE2* repair fragment
10357	GTTTATTGGATGGCAATCTCG	Construction of *O. parapolymorpha ADE2* repair fragment
10737	AATTGCATCTCTTTGTGATGTC	Sanger sequencing of *K. lactis ADE2* disruptions

SeqADE2	CTGCAACTGCTTGTTCAGCC	Sanger sequencing of *K. marxianus ADE2* disruptions
10378	CCAATTACAAGTACTACTTCGAG	Sanger sequencing of *O. polymorpha ADE2* disruptions fw
10379	CTAGCTCCTTGGTGAAAGG	Sanger sequencing of *O. polymorpha ADE2* disruptions rv
10386	ACAAGTACTACTTCGAGGAC	Sanger sequencing of *O. parapolymorpha ADE2* disruptions fw
10387	CTAGCTCCTTGGTAAAGGG	Sanger sequencing of *O. parapolymorpha ADE2* disruptions rv
10751	GGACGCCTGCTTAGACTTG	Sanger sequencing of *O. parapolymorpha KU80* disruptions fw
10752	AGCACGGTATATTCGCACAG	Sanger sequencing of *O. parapolymorpha KU80* disruptions rv
12257	CACCATGGTCGGAAGAACC	Sanger sequencing of *O. parapolymorpha YNR1* disruptions fw
12266	ATGTAATTCCTCACGAACTTTGG	Sanger sequencing of *O. parapolymorpha YNR1* disruptions rv
10909	TCTTCGTCGCCATTTATTGTTGAG	Diagnosis of *K. lactis ADE2* deletions fw
10910	CTATTGCGGTTCGCTCTTCC	Diagnosis of *K. lactis ADE2* deletions rv
10911	ATTCGCCGAATCTGACGTG	Diagnosis of *K. marxianus ADE2* deletions fw
10912	TGGTGTGCAGACGGATAAAC	Diagnosis of *K. marxianus ADE2* deletions rv
10380	AGGTGCTCAAACACAAAGAG	Diagnosis of *O. polymorpha ADE2* deletions fw
10381	TCGTATCTCGTAAGTTGATTTAGG	Diagnosis of *O. polymorpha ADE2* deletions rv
10915	CCGTCTGAACGGAATGATGTC	Diagnosis and Sanger sequencing of *O. parapolymorpha ADE2* deletions fw
10916	CCCTCAACTGCAGACACATAG	Diagnosis and Sanger sequencing of *O. parapolymorpha ADE2* deletions rv

DNA fragments obtained by PCR were separated by gel electrophoresis. Gel purification was carried out using the Zymoclean Gel DNA Recovery Kit (Zymo Research, Irvine, CA, USA). PCR purification was performed using either the GenElute PCR Clean-Up Kit (Sigma-Aldrich) or GeneJET PCR purification kit (Thermo Fisher Scientific). Gibson assembly was done using the NEBuilder HiFi DNA Assembly Master Mix (New England Biolabs, Ipswich, MA, USA) according to the manufacturer's recommendations. Restriction digest with PdmI and SmaI was performed using FastDigest enzymes (Thermo Fisher Scientific), according to the manufacturer's instructions. *E. coli* strain XL1-blue was used for plasmid transformation, amplification and storage. Plasmids were isolated from *E. coli* with the GenElute Plasmid Miniprep Kit (Sigma-Aldrich).

Whole-genome sequencing of *O. parapolymorpha* IMD001. Genomic DNA of *O. parapolymorpha* IMD001 was isolated with the QIAGEN Genomic-tip 100/G kit (Qiagen, Hilden, Germany) from a stationary-phase overnight shake-flask culture grown on YPD medium, according to the manufacturer's instructions. Genomic DNA was quantified using a Qubit® 2.0 Fluorometer (Thermo Fisher Scientific). 10 mg of genomic DNA was sequenced by Novogene Bioinformatics Technology (Yuen Long, Hong Kong, China) on a HiSeq 2500 (Illumina, San Diego, CA, USA) with 150-bp paired-end reads using TruSeq PCR-free library preparation (Illumina). In order to verify complete absence of plasmid pUDP046, sequencing reads were mapped to the sequence of pUDP046 and to the genome of *O. parapolymorpha* CBS 11895 (accession GCF_000187245.1) (Ravin *et al.* 2013) using the Burrows-Wheeler Alignment tool and further processed using SAMtools (Li and Durbin 2010, Li *et al.* 2009). The absence of sequences from pUDP046 was confirmed by visualizing the generated .bam files in the Integrative Genomics Viewer software (Robinson *et al.* 2011). Sequencing data are available at NCBI (https://www.ncbi.nlm.nih.gov/) under BioProject PRJNA413643.

Results

pUDP: a plasmid-based wide-host-range yeast CRISPR/Cas9 system. The pUDP system was designed to enable Cas9-mediated genome editing in different yeast species by simple transformation with a single plasmid. To this end, DNA parts encoding the plasmid origin of replication, *cas9* expression cassette, selection marker and gRNA expression cassette were chosen to function in a wide range of yeast species (**Figure 2.1**). To ensure replication of pUDP plasmids, the pangenomic yeast origin of replication panARS(OPT) was used. This origin of replication, which was inspired by a *K. lactis* chromosomal ARS, has been shown to function in at least ten different species of ascomycetous yeasts, including the species used in this study (Liachko and Dunham 2014). The *cas9*$^{D147Y\ P411T}$ nuclease variant (Bao *et al.* 2015) was expressed under control of the *TEF1* promoter from *Arxula adeninivorans* (*Blastobotrys adeninivorans*), which enabled strong constitutive expression in various yeasts (Steinborn *et al.* 2006, Terentiev *et al.* 2004). Similarly, the *K. pneumoniae hph* gene encoding hygromycin B phosphotransferase, which conferred hygromycin resistance in a wide range of microorganisms, was expressed under control of the *Ashbya gossypii* (*Eremothecium gossypii*) *TEF1* promoter (Wach *et al.* 1994) that showed activity in a range of yeasts (Kim *et al.* 2010, Mazzoni *et al.* 2005).

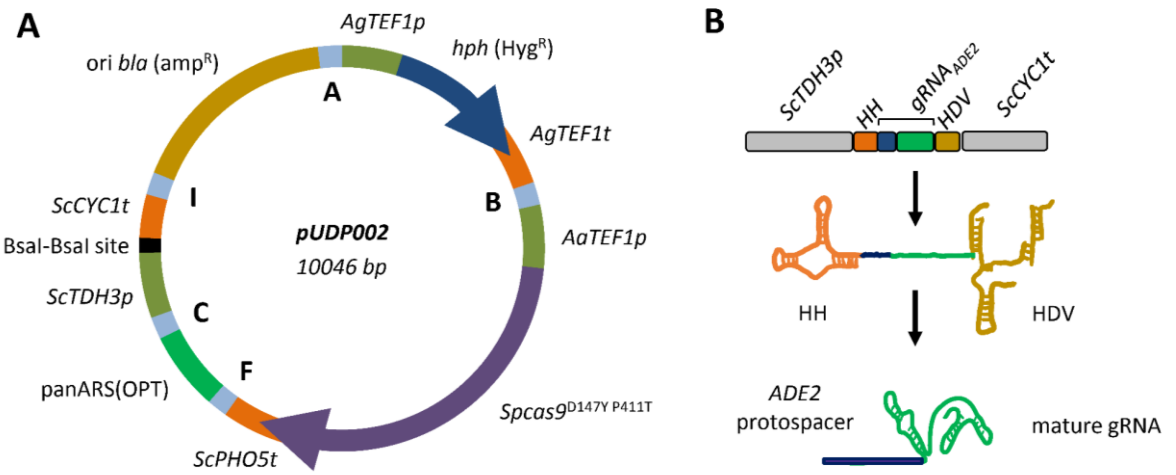


Figure 2.1: Components of the pUDP genome editing system. (A) Map of pUDP002 (Addgene plasmid #103872), a wide-host-range gRNA entry plasmid. pUDP002 is composed of a *hph* (Hyg^R) selection marker cassette under control of the *TEF1* promoter from *Ashbya gossypii* conferring hygromycin resistance, *Spcas9*$^{D147Y\ P411T}$ under control of the *TEF1* promoter from *Arxula adeninivorans*, the pangenomic yeast replication origin panARS(OPT), a BsaI cloning site for entry of gRNA constructs, and a pBR322-derived *E. coli* origin and *bla* gene for ampicillin selection. A, B, F, C and I indicate 60 bp synthetic homologous recombination sequences used for 'Gibson' assembly of the plasmid. **(B)** Representation of the ribozyme-flanked gRNA expression cassette design. Guide RNAs (represented by gRNA$_{ADE2}$) were flanked on their 5′ by a hammerhead ribozyme (HH represented in orange) and on their 3′ by a hepatitis delta virus ribozyme (HDV represented in bronze). When integrated into pUDP002, this construct is under control of the RNA polymerase II promoter *TDH3* and the *CYC1* terminator from *S. cerevisiae*. Upon ribozyme self-cleavage, a mature gRNA comprised of the guiding protospacer (in blue) and the structural gRNA fragment (in green) is released.

To avoid RNA modifications that interfere with biological function, transcription of gRNAs for genome editing is commonly controlled by RNA polymerase III promoters. By analogy with expression of gRNA in mammalian cells (Cong *et al.* 2013, Mali *et al.* 2013), the RNA polymerase III SNR52 promoter was used in *S. cerevisiae* (DiCarlo *et al.* 2013, Mans *et al.* 2015), but accurate annotation and characterization of this promoter might not be available for other yeast species. Therefore, self-processing ribozyme-flanked gRNAs were expressed from an RNA polymerase II promoter (Gao and Zhao 2014, Ryan *et al.* 2014). This concept has already been successfully applied in several different organisms (Gao *et al.* 2016, Gorter de Vries *et al.* 2017, Weninger *et al.* 2016). In this system, the gRNA is flanked by a hammerhead and hepatitis delta virus ribozyme at its 5' and 3' end, respectively, resulting in precise release of mature gRNA after self-cleavage. To ensure sufficient gRNA transcription, the strong glycolytic *TDH3* promoter from *S. cerevisiae* was used. These elements were combined in plasmid pUDP002, which could subsequently serve as entry plasmid to insert any desired gRNA. To streamline integration of ribozyme-flanked gRNAs, pUDP002 contained two different BsaI restriction sites between *ScTDH3p* and *ScCYC1t*, enabling directional insertion of synthetic BsaI-flanked gRNA constructs (Gorter de Vries *et al.* 2017).

To test the genome-editing efficiency of the pUDP system, the *ADE2* gene was targeted in all four non-conventional yeasts used in this study: *K. lactis*, *K. marxianus*, *O. polymorpha* and *O. parapolymorpha*. The *ADE2* gene encodes a phosphoribosyl-aminoimidazole carboxylase, also referred to as AIR carboxylase, involved in adenine biosynthesis. Besides causing adenine auxotrophy, loss-of-function mutations in *ADE2* result in a red-color phenotype due to accumulation of the oxidized form of 5-amino imidazole ribonucleotide (AIR). This allows detection of *ADE2* targeting by simple visual inspection of transformation plates (Roman 1956). Therefore, gRNAs were designed targeting *ADE2* in each species based on available genome data (**Table 2.3**), and gRNA-harboring plasmids pUDP025, pUDP082 and pUDP013 for deletion of *ADE2* in *K. lactis*, *K. marxianus* and *O. polymorpha*/*O. parapolymorpha*, respectively, were constructed.

Efficient CRISPR/Cas9 targeting enables gene deletion in *Kluyveromyces* species. To test the effectiveness of the pUDP system in *K. lactis*, gRNA$_{KlADE2}$ was inserted into pUDP002 and the resulting plasmid pUDP025 was used to target *ADE2* in *K. lactis* CBS 2359. Transformation with pUDP025 without a repair DNA fragment yielded a total of 35 transformants of which 19 (54%) exhibited a red Ade$^-$ phenotype. In the presence of a 962-bp repair fragment which targeted the *ADE2* promoter and terminator (**Figure 2.2A**), the transformation of pUDP025 yielded 26 red transformants out of a total 27 (96%) (**Figure 2.2B**).

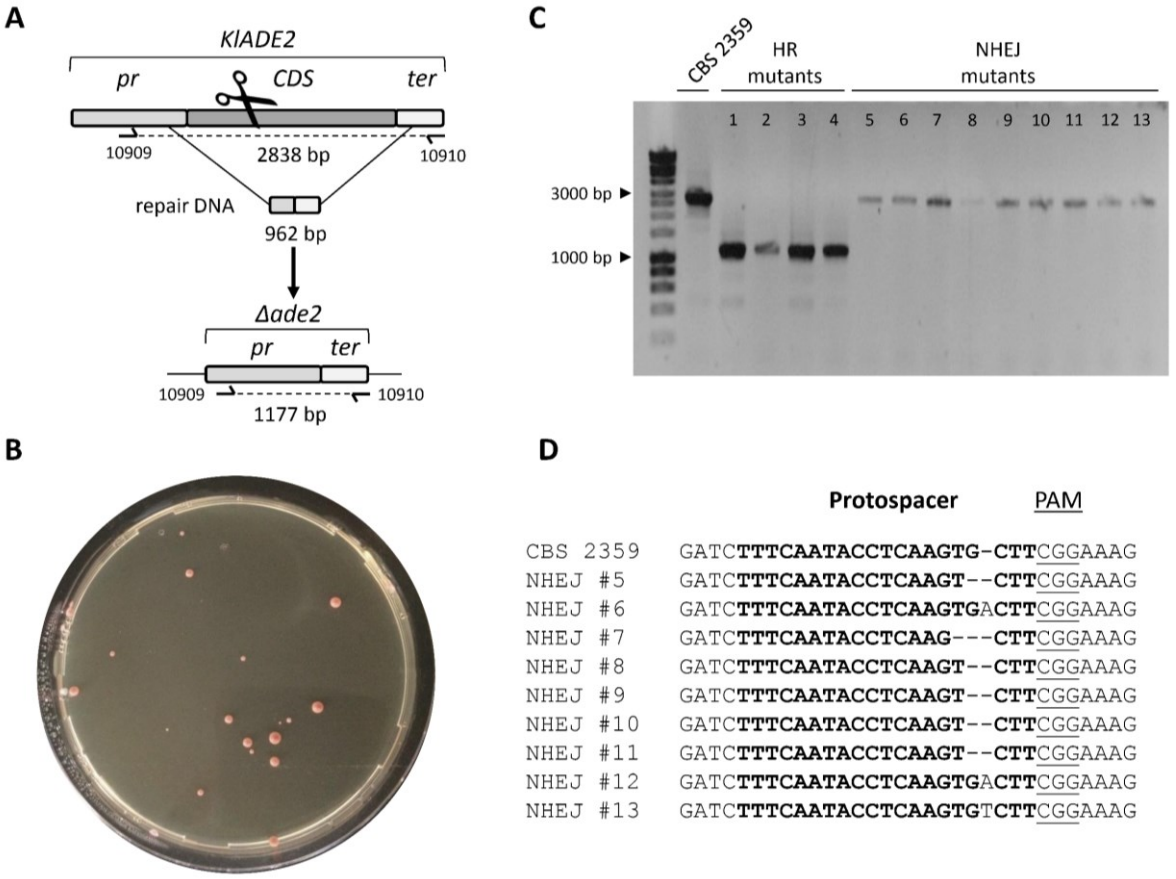


Figure 2.2: Efficient gRNA targeting in *K. lactis* CBS 2359 enables marker-free gene deletion. (A) Schematic representation of *ADE2* editing upon transformation of CBS 2359 with pUDP025 (gRNA$_{KlADE2}$) and a repair DNA fragment. The primers for diagnostic PCR of transformants are indicated. **(B)** Colony morphology of CBS 2359 upon transformation with pUDP025 and a marker-free 962-bp repair fragment. **(C)** Diagnosis of 13 randomly picked red Ade$^-$ transformants of CBS 2359 upon transformation with pUDP025 and a 962-bp marker-free repair fragment. Four transformants (HR mutants 1–4) showed a PCR product of 1177 bp corresponding to the deleted allele. The control labeled CBS 2359 and nine transformants (NHEJ mutants 5–13) showed a PCR product of 2838 bp corresponding to the wild-type allele. **(D)** Sanger sequencing results of purified PCR fragments from nine Ade$^-$ mutants (corresponding to mutants 5–13 in panel C) derived from the transformation of CBS 2359 with pUDP025 and repair fragment.

In comparison, transformation of the backbone plasmid pUDP002, which does not express a gRNA, generated a number of transformants that was ca. 35-fold higher, none of which displayed a red phenotype. This difference in transformation efficiency already provided information about the quality of the gRNA and the editing. Unless repaired a chromosomal DSB should be lethal, therefore this 35-fold drop in transformation efficiency would represent the fraction of transformants that did not successfully repair the CRISPR-Cas9 induced break and subsequently died from it. Out of the 26 Ade$^-$ transformants obtained after co-transformation with pUDP025 and repair fragment, 13 red colonies were randomly picked and analyzed by diagnostic PCR, revealing two distinct groups. Nine colonies exhibited a normally-sized *ADE2* locus, while four colonies (31%) showed a PCR-fragment size compatible with complete deletion of the *KlADE2* ORF (**Figure 2.2C**). Subsequent Sanger sequencing of the

amplified fragments derived from the nine clones exhibiting an *ADE2* wild-type size band identified the presence of indels at the targeted site. These indels reflected imperfect repair of the Cas9-induced DSB via NHEJ, resulting in introduction of loss-of-function mutations and disruption of the Cas9 target site (**Figure 2.2D**). These results demonstrated that the pUDP system could be used for efficient targeting of the *ADE2* gene in *K. lactis* CBS 2359, resulting in repair by either HR or NHEJ DNA repair mechanisms.

The pUDP system was similarly tested in *K. marxianus* by targeting *ADE2* in the haploid strain NBRC 1777 and the diploid strain CBS 397, using plasmid pUDP082 expressing gRNA$_{KmADE2}$. The transformation of *K. marxianus* NBRC 1777 with pUDP082, without and with a 959-bp repair DNA fragment resulted in 13 and 30 transformants, respectively. Of these, 13 out of 13 and 29 out of 30 transformants were red, indicating successful disruption of *ADE2* in both cases. To determine which repair mechanism contributed to the repair of the Cas9-induced DSB in the presence of the repair fragment, 13 randomly picked red transformants were subjected to diagnostic PCR. Similar to the results obtained in *K. lactis*, two groups of transformants were identified. Three transformants (24%) showed a genotype corresponding to the repair of the DSB with the repair fragment (**Figure 2.3A-B**), while the remaining transformants exhibited a PCR result that was undistinguishable from that of the NBRC 1777 *ADE2* wild-type locus. Sanger sequencing of the Cas9 target site again revealed the presence of indels, leading to nonsense mutations in the *ADE2* coding sequence (**Figure 2.3C**).

When the diploid strain CBS 397 was transformed with pUDP082 without repair DNA fragment, 106 out of 133 (80%) of the colonies were red, while 117 out of 143 (82%) colonies showed the red phenotype when a 959-bp repair fragment was provided. In comparison, a control transformation with pUDP002 yielded 262 transformants without any red phenotypes. From the transformation with pUDP082 and repair fragment, ten red colonies were randomly picked and analyzed by diagnostic PCR, resulting in PCR products that were identical in size to the native *ADE2* locus in all cases. Sanger sequencing of the same mutants confirmed the link between the red-colored phenotype and occurrence of small indels at the Cas9 cut site, which likely introduced loss-of-function mutations in *ADE2* (**Figure 2.3D**). Interestingly, Sanger sequencing also indicated that the editing occurred identically at both *KmADE2* alleles as a clear and continuous sequence was observed over the cut site. These results demonstrated that the *ADE2* gene can be efficiently targeted by the pUDP system in both *K. marxianus* strains, with DNA repair mediated by HR or NHEJ mechanisms in the haploid strain NBRC 1777 and NHEJ in the diploid strain CBS 397.

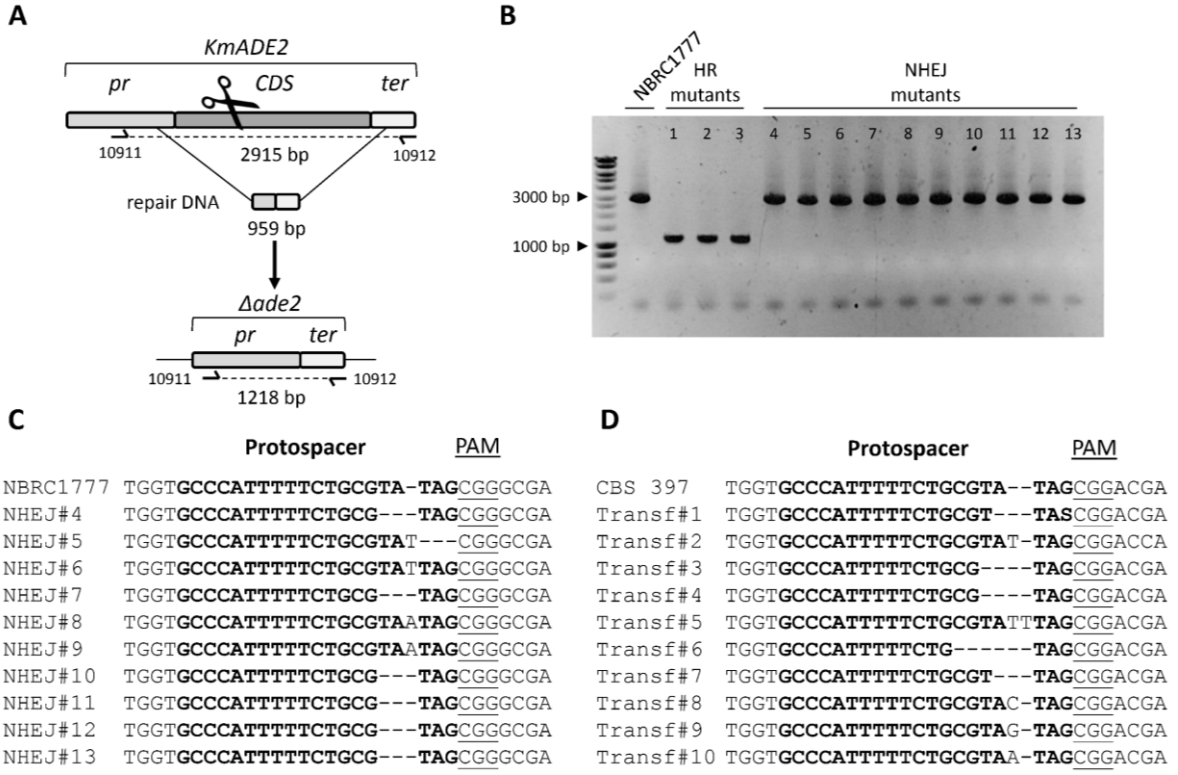


Figure 2.3: Efficient gRNA targeting enables marker-free gene deletion in haploid *K. marxianus* NBRC 1777 and gene disruption in diploid *K. marxianus* CBS 397. (A) Schematic representation of the *ADE2* editing upon transformation of NBRC 1777 with pUDP082 (gRNA$_{KmADE2}$) and a repair DNA fragment. The primers for diagnostic PCR of transformants are indicated. **(B)** Diagnosis of 13 randomly picked red Ade⁻ transformants of NBRC 1777 upon transformation with pUDP082 and a 959-bp marker-free repair fragment. Three transformants (HR mutants 1–3) showed a PCR product of 1218 bp corresponding to the deleted allele. The control labeled NBRC 1777 and 10 transformants (NHEJ mutants 4–13) showed a PCR product of 2915 bp corresponding to the wild-type allele. **(C)** Sanger sequencing results of purified PCR fragments from 10 Ade⁻ mutants (corresponding to mutants 4–13 in panel B) derived from the transformation of NBRC 1777 with pUDP082 and repair fragment. **(D)** Sanger sequencing results of purified PCR fragments of 10 randomly picked red Ade⁻ mutants derived from the transformation of CBS 397 with pUDP082 and repair fragment.

CRISPR/Cas9 editing enables gene deletion in *Ogataea* species. To test the effectiveness of the pUDP system in *O. parapolymorpha*, plasmid pUDP013 harboring gRNA$_{OpADE2}$ was used to transform strain CBS 11895 (DL-1). Transformations were performed with or without a 960-bp repair DNA fragment designed to mediate HR at the promoter and terminator regions of the *OpADE2* gene (**Figure 2.4A**). In presence of the repair fragment, an average of 298 ± 50 colonies per transformation were obtained, while a lower average number of 64 ± 9 colonies were counted after transformation without the repair fragment. In contrast to the situation in *Kluyveromyces* species, none of these colonies exhibited a red Ade⁻ phenotype, although the ability of the transformants to grow on selective medium indicated that the plasmid was present. The heterologous origin of genetic parts on pUDP002 might result in a suboptimal expression of the necessary components of the CRISPR-Cas9 machinery in *O. parapolymorpha*. To allow more time for expression of Cas9 and gRNA, two randomly picked transformants (CBS 11895 + pUDP013) from independent experiments were

incubated for a longer period in selective YPD medium. These prolonged liquid incubation cultures were sampled after 48, 96 and 192 h, and samples taken at each of these time points were plated on selective YPD medium. Already after 48 h of incubation, 31 ± 7% of the plated colonies exhibited the red Ade⁻ phenotype. This fraction increased to 61 ± 2% and 63 ± 2% after 96 and 192 h of incubation, respectively (**Figure 2.4B**). From a plate obtained after 96 h of liquid incubation, ten red colonies were randomly picked and subjected to diagnostic PCR. All ten transformants showed a band compatible with the size of the wild-type *ADE2* locus (**Figure 2.4C**). As previously found in *K. lactis* and *K. marxianus*, Sanger sequencing of the *ADE2* gene in these ten transformants (NHEJ #1 to #10) confirmed the presence of short indels at the Cas9 target site, which introduced nonsense mutations within *ADE2* (**Figure 2.4D**).

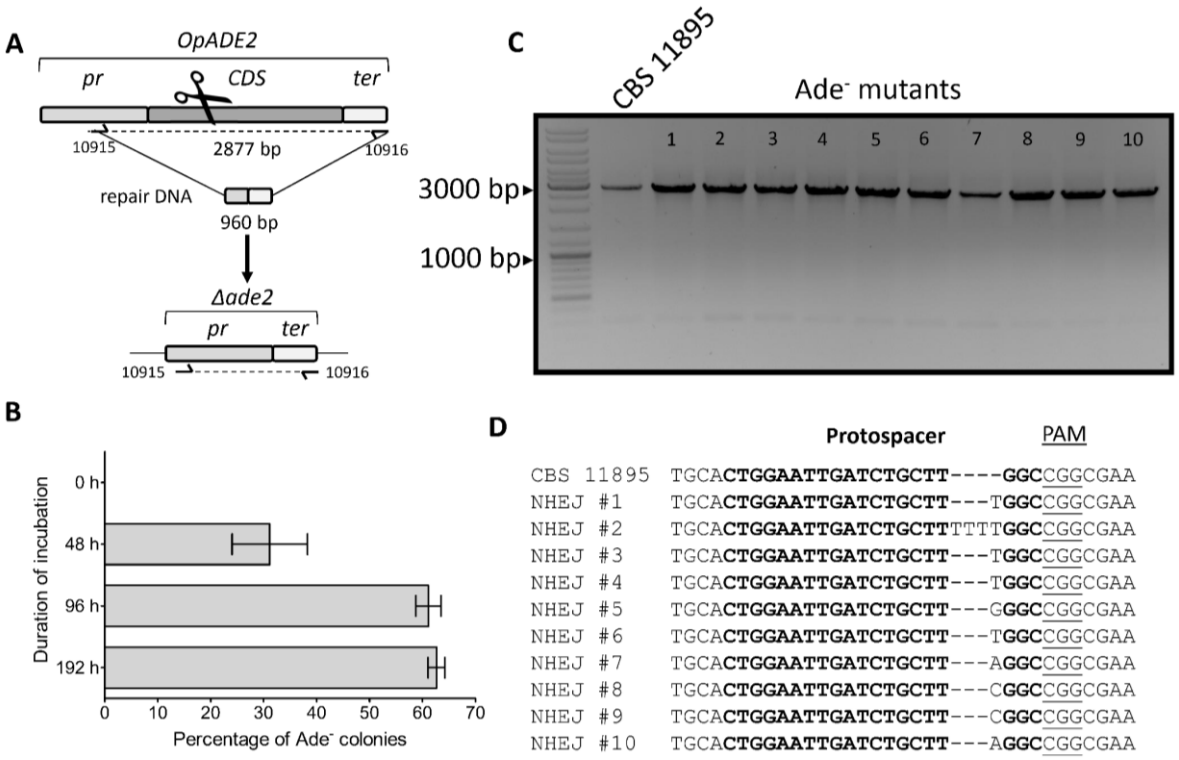


Figure 2.4: Prolonged liquid incubation enables gene disruption in *O. parapolymorpha* CBS 11895 (DL-1). **(A)** Schematic representation of the potential *ADE2* editing upon transformation of CBS 11895 with pUDP013 (gRNA*OpADE2*) and a repair DNA fragment. **(B)** Percentage of red Ade⁻ colonies observed based on phenotypic screening of CBS 11895 + pUDP013 (gRNA$_{OpADE2}$) transformants when plated directly after transformation (0 h), or after prolonged liquid incubation of established transformants in selective medium (for 48, 96 and 192 h) and subsequent plating. Mean and standard deviation were calculated with a total of 4411, 4710 and 4301 colonies (obtained after 48, 96 and 192 h, respectively) from two liquid incubation cultures started with single CBS 11895 transformants which were obtained in independent transformations. No red Ade⁻ mutants were observed when cells were plated directly after transformation. **(C)** Diagnosis of 10 randomly picked red Ade⁻ colonies of CBS 11895 after transformation with pUDP013 and 96 h of subsequent liquid incubation under selective conditions. All 10 transformants (NHEJ mutants 1–10) and the control labeled CBS 11895 showed a PCR product of 2877 bp corresponding to the wild-type allele. **(D)** Sanger sequencing results of purified PCR fragments from 10 red Ade⁻ colonies (corresponding to mutants 1–10 in panel B) derived from the transformation of CBS 11895 with pUDP013 and 96 h of subsequent liquid incubation.

The presence of the repair fragment did not change this outcome: prolonged liquid incubation cultures that were started from colonies obtained by co-transformation of pUDP013 and the repair fragment gave rise to identical proportions of Ade$^-$ mutants as those described above. For the closely related species *O. polymorpha*, the obtained results were qualitatively comparable. Transformation of strain *O. polymorpha* CBS 4732 with pUDP013 did not result in colonies with a red phenotype, neither in the presence nor in the absence of a 960-bp repair DNA fragment. Also, no mutations in *ADE2* were detected as evaluated by Sanger sequencing of nine randomly picked (white) colonies from a transformation without the repair fragment. When two independent transformants (CBS 4732 + pUDP013) were incubated in liquid medium, plating after 192 h yielded 9 ± 1% of red colonies. These results indicated that the pUDP system promoted less efficient genome editing in the *O. polymorpha* strain CBS 4732 than in *O. parapolymorpha* CBS 11895, although the system could be utilized for NHEJ-mediated disruption of the *ADE2* gene in both yeasts.

In contrast to *S. cerevisiae* (Baudin *et al.* 1993) and *Kluyveromyces* species (this study), the HR-mediated DNA repair mechanism in *Ogataea* species is relatively inefficient (Klinner and Schäfer 2004). The removal of *KU80* in *O. polymorpha* has been reported to result in a strong reduction of NHEJ in return favoring the occurrence of HR-mediated DNA repair (Saraya *et al.* 2012). To delete *KU80* in *O. parapolymorpha*, plasmid pUDP046 harboring gRNA$_{OpKU80}$ was constructed and used to transform strain CBS 11895 (**Figure 2.5A**). Following the liquid incubation procedure established for disruption of *OpADE2*, a single transformant was picked, incubated in selective YPD medium for 96 h and plated on selective YPD plates. Of eight colonies randomly picked and subjected to Sanger sequencing of the *KU80* locus, five (NHEJ #3 to #7) contained a single adenine nucleotide inserted at the Cas9 targeting site between position 340 and 341 of the *OpKU80* coding sequencing (A340AA), resulting in a loss-of-function mutation (**Figure 2.5B**). One of these transformants was isolated by restreaking three times on non-selective medium, and renamed IMD001 (*Opku80*^{A340AA}).

While strain IMD001 was not able to grow on selective medium indicating loss of pUDP046, complete curation of the plasmid had to be verified in order to use IMD001 for subsequent genetic interventions with pUDP plasmids. Circular plasmids with limited homology to the nuclear DNA have been reported to integrate into the genome of *Ogataea* (*Hansenula*) yeasts, despite the presence of an origin of replication on the plasmid (Kunze *et al.* 2009). In particular, upon cultivation under selective conditions, plasmids were found to have integrated into the nuclear genome with high variations in copy number in strain CBS 11895 (DL-1) (Kang *et al.* 2001). To verify that no pUDP046 plasmid sequences had recombined into the chromosomal DNA of strain IMD001, its genome was sequenced using Illumina 150-bp paired-end short reads that were mapped onto a reference comprised of an *O. parapolymorpha* CBS 11895 genome

assembly (accession number: GCF_000187245.1) and the sequence of pUDP046. No reads were found to map onto the pUDP046 sequence, thus confirming the absence of unwanted integration of the transformed plasmid into the nuclear genome and the suitability of *O. parapolymorpha* IMD001 as a host strain for further genetic engineering.

To test the impact of the *KU80* disruption on the pUDP system in *O. parapolymorpha*, plasmid pUDP013 (gRNA$_{OpADE2}$) was used to transform strain IMD001 together with the 960-bp repair DNA fragment. As previously observed for the wild-type strain *O. parapolymorpha* CBS 11895, no red Ade$^-$ transformants were observed on the transformation plates. In an attempt to improve occurrence of homology-mediated repair, a liquid incubation culture was directly inoculated with the transformation mix, grown for 48 h and then plated on selective YPD medium. On these plates, only seven out of ~1900 colonies exhibited a red Ade$^-$ phenotype, representing a <1% targeting efficiency.

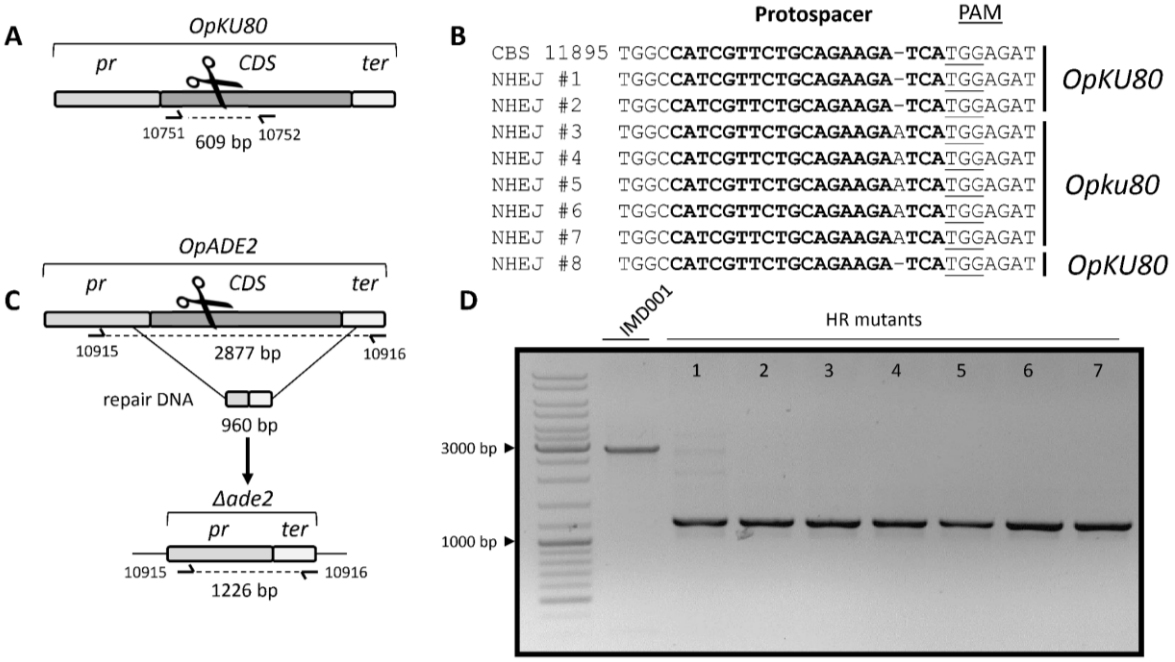


Figure 2.5: Construction of IMD001, an *O. parapolymorpha ku80* strain that enables low efficiency gene deletion. (A) Schematic representation of the *OpKU80* editing upon transformation of CBS11895 with pUDP046 (gRNA$_{OpKU80}$). The primers for diagnostic PCR and sequencing of transformants are indicated. **(B)** Sanger sequencing results of purified PCR fragments from eight randomly picked colonies derived from the transformation of CBS 11895 with pUDP046 (gRNA$_{OpKU80}$) and 96 h of subsequent liquid incubation. The transformants NHEJ#1, NHEJ#2 and NHEJ#8 displayed a wild-type sequence, while mutants NHEJ #3 to #7 included frameshift mutations **(C)** Schematic representation of the *ADE2* editing upon transformation of *O. parapolymorpha* IMD001 (CBS 11895 *ku80*) with pUDP013 (gRNA$_{OpADE2}$) and a repair DNA fragment. The primers for the diagnostic PCR of transformants are indicated. **(D)** Diagnosis of all seven red Ade$^-$ colonies obtained (from a total of ca. 1900 colonies) upon transformation of IMD001 with pUDP013 and a 960-bp marker-repair fragment, and subsequent liquid incubation in selective conditions for 48 h, started directly from the transformation recovery culture. All mutants showed a PCR product of 1226 bp corresponding to the deleted allele. The control labeled IMD001 showed a PCR product of 2915 bp corresponding to the wild-type allele.

Diagnostic PCR performed on these seven transformants revealed that all harbored a deleted *ADE2* version compatible with HR-mediated repair of the CRISPR-Cas9-induced DSB (**Figure 2.5C-D**). One of the obtained mutants was further analyzed by Sanger sequencing, which confirmed scarless integration of the repair fragment. Longer liquid incubation times of up to 192 h did not increase the incidence of red colonies after plating. These results demonstrated that the *KU80* disruption abolished NHEJ-mediated DSB repair in *O. parapolymorpha* IMD001, and that the CBS 11895 strain background possesses a basal HR-mediated DNA repair activity that can be utilized for precise genome engineering. However, the low incidence of HR-mediated repair precludes efficient genome editing unless the desired phenotype is easily screenable.

Multiplexed gene targeting in *O. parapolymorpha* by expression of double ribozyme-flanked gRNA array. As recently demonstrated, HH-HDV ribozyme-flanked gRNAs can be concatenated into polycistronic arrays enabling multiplexing gene editing in *S. pastorianus* (Gorter de Vries *et al.* 2017) and transcriptional interference in *S. cerevisiae* (Deaner *et al.* 2017). To explore this possibility in *O. parapolymorpha*, a gRNA expression plasmid pUDP123 carrying spacers targeting *OpADE2* and *OpYNR1* was designed. The additionally targeted *YNR1* gene encodes a nitrate reductase, which is involved in utilization of nitrate as N-source (Brito *et al.* 1996, Navarro *et al.* 2003). *Ogataea* strains harboring a non-functional nitrate reductase are unable to grow on media with nitrate as sole N-source.

A tandem array of [HH-gRNA-HDV] targeting *OpADE2* and *OpYNR1* spaced with a 20-bp linker and expressed under control of the *ScTDH3* promoter as previously described (Gorter de Vries *et al.* 2017) was carried by the recombinant plasmid pUDP123 (*hph cas9 ScTDH3p*-HH-gRNA$_{OpADE2}$-HDV-HH-gRNA$_{OpYNR1}$-HDV-*ScCYC1t*) that was transformed in *O. parapolymorpha* CBS 11895 (**Figure 2.6A**). Following the liquid incubation procedure established for disruption of *OpADE2* and *OpKU80*, a transformant was picked, incubated in selective YPD medium for 96 h and then plated onto selective YPD medium. Of the resulting colonies, 32% were exhibiting a red Ade$^-$ phenotype. To verify whether these Ade$^-$ mutants were able or not to grow on nitrate, 94 red transformants were randomly picked and grown on non-selective medium (YPD). After full growth, 10 µl of cell suspension was replicated on non-selective (SM) and selective (SMN) media supplemented with 15 mg L^{-1} adenine to complement the Ade$^-$ phenotype. About 10% of the red-phenotype transformants gradually returned white. While one cannot exclude a reversion of the mutation, the pattern observed would suggest that the replicated red colonies were not pure. Growth of the transformant on SMN was disturbed by the addition of adenine which might also serve as nitrogen source as *O. parapolymorpha* is equipped with an adenine deaminase. However, since the adenine concentration remained ~300-fold lower than the nitrate

concentration, a significant difference between Nit$^+$ and Nit$^-$ strains should be noticeable. Indeed out of the 94 red transformants, 17 showed a strongly reduced growth on SMN supplemented with adenine when compared to the control CBS 11895, and five of these potential double mutants were further analyzed by Sanger sequencing. While all five were confirmed to harbor a frameshift around the CRISPR cut site in *OpADE2*, only three transformants were concomitantly exhibiting a frameshift in *OpYNR1* (**Figure 2.6C**). Transformant #2 was isolated by restreaking three times on non-selective medium, and renamed IMD034 (*Opade2*^{C120CA} *Opynr1*^{G397GT}). Assuming that all other red clones that were also growing slowly on SMN medium shared this genotype, double editing would have occurred at a rate of 2%–5%.

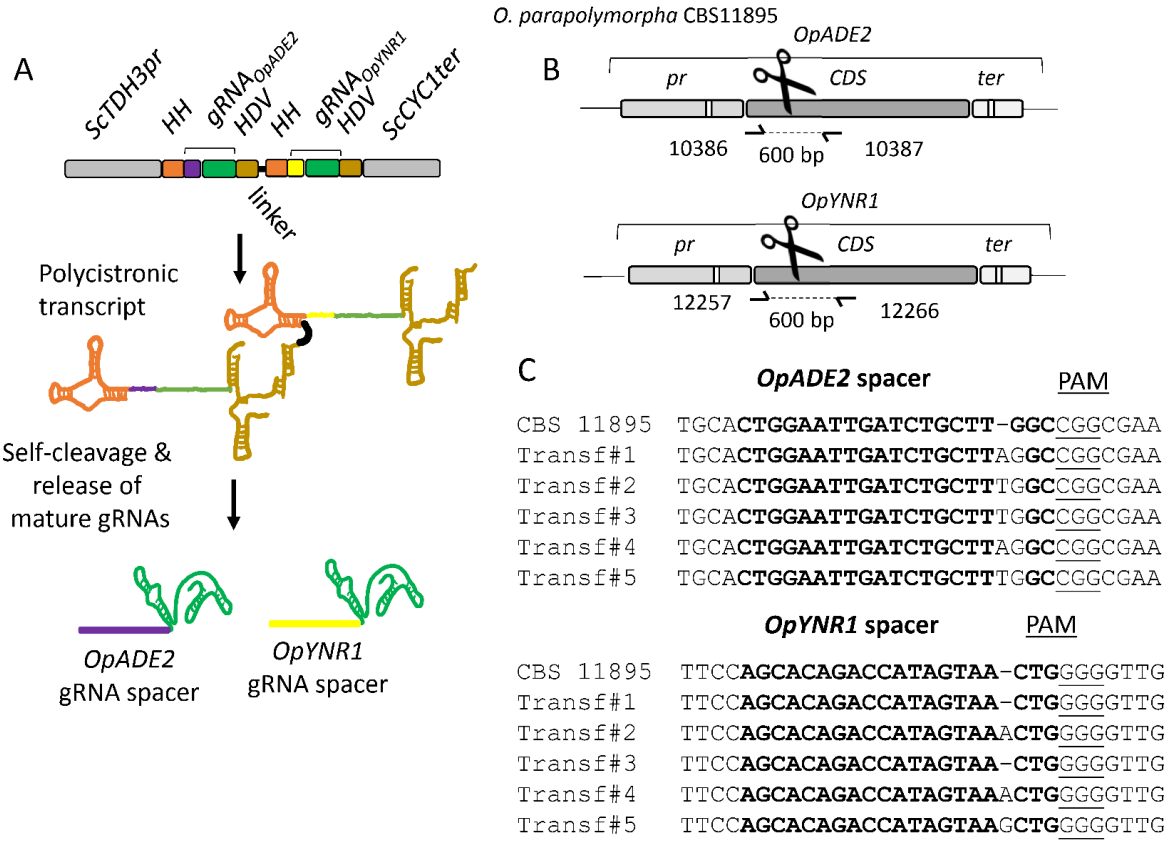


Figure 2.6: Simultaneous deletion of *OpADE2* and *OpYNR1* alleles using a single ribozyme flanked gRNA array in *O. parapolymorpha* CBS11895. (A) Representation of the gRNA array expression cassette in pUDP123. The dual gRNA array was under the control of the RNA polymerase II promoter *ScTDH3* and *ScCYC1* terminator. Each gRNA was flanked on its 5′ by a hammerhead ribozyme (HH represented in orange) and on its 3′ by a hepatitis delta virus (HDV represented in bronze) ribozyme which were separated by a 20-bp linker. Upon ribozyme self-cleavage, the mature gRNAs are released. The *OpADE2* guiding spacer (in purple), the *OpYNR1* guiding spacer (in yellow) and the constant structural gRNA fragment (in green) are indicated. **(B)** Schematic representation of the *OpADE2* and *OpYNR1* loci of CBS11895. The primers for the validation of transformants are indicated. **(C)** Sanger sequencing results of *OpADE2* and *OpYNR1* editing site of five randomly picked red Ade$^-$ mutants that have lost ability to grow on nitrate. Transformant labeled Transf#2 was renamed IMD034.

Discussion

The pUDP system, which combines generic DNA parts with expression of self-processing gRNAs driven by RNA polymerase II, was shown to enable genome editing in different yeasts belonging to the *Saccharomycotina* subphylum. The pUDP002 entry plasmid only required insertion of a functional gRNA to enable successful single deletion or disruption of the *ADE2* gene in four different species of *Kluyveromyces* and *Ogataea* yeasts (**Figure 2.7**) but also double editing at two different chromosomal loci in *O. parapolymorpha*. The panARS(OPT)-harboring pUDP plasmids can be cured simply by growth in non-selective medium (Liachko and Dunham 2014), as illustrated by whole-genome sequencing of engineered *O. parapolymorpha* strain IMD001, enabling sequential rounds of genetic modifications with the same system in rapid succession.

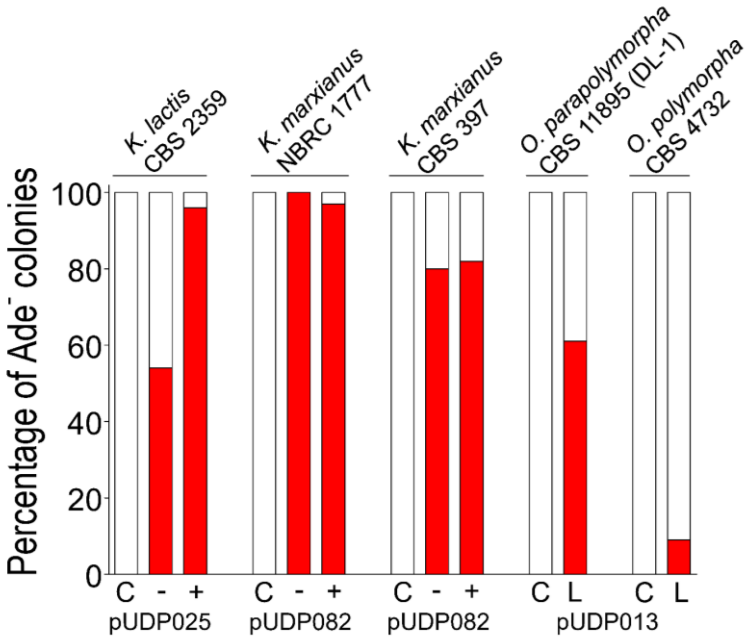


Figure 2.7: Wide-host-range applicability of the pUDP CRISPR/Cas9 system. The depicted data summarize the *ADE2* targeting efficiency (red colonies/total colonies) of the pUDP system in the four yeast species used in this study. The pUDP plasmids differ only in their gRNA target specificity: pUDP025, pUDP082 and pUDP013 harbor gRNA$_{KlADE2}$, gRNA$_{KmADE2}$ and gRNA$_{OpADE2}$, respectively. The results shown for *Kluyveromyces* species were obtained directly on transformation plates either with (+) or without (−) co-transformation of a ca. 960-bp marker-free repair DNA fragment. The results for *O. parapolymorpha* and *O. polymorpha* were obtained after 96 and 192 h of prolonged liquid incubation (L), respectively, started from colonies that were transformed without a repair fragment. Control transformations (C) with pUDP002 did not result in the occurrence of Ade⁻ mutants.

In *Kluyveromyces* species, near-perfect targeting (≥96%) and HR-mediated repair DNA integration occurred at practicable rates in *K. lactis* CBS 2359 (31%) and the haploid *K. marxianus* NBCR 1777 (24%), while highly efficient gene disruption was achieved in the diploid *K. marxianus* strain CBS 397. In *Ogataea* species, mutants could be obtained after prolonged liquid incubation, allowing relatively straightforward and marker-free

disruption of genes in the two industrially relevant *O. (para)polymorpha* strains CBS 4732 and CBS 11895 (DL-1). The observed delay in targeting activity suggests that the heterologous origin of functional parts of the pUDP system might reflect suboptimal expression of gRNA and/or Cas9 in *Ogataea* species. The expression levels of gRNA and Cas9 may have narrow optima for efficient genome editing, as recently demonstrated in *P. pastoris*, a related methylotrophic yeast (Weninger *et al.* 2016). While the *A. adeninivorans TEF1* promoter employed for Cas9 expression in this study has been demonstrated to enable strong constitutive expression in *O. polymorpha* (Terentiev *et al.* 2004), optimization of the gRNA expression presently under the control of the *S. cerevisiae TDH3* promoter might be envisaged to eliminate the observed delay in the occurrence of mutants.

Recently, species-specific CRISPR-Cas9 systems have been published for a range of non-conventional yeasts including *K. lactis*, *K. marxianus* and *O. polymorpha* (Horwitz *et al.* 2015, Löbs *et al.* 2017a, Nambu-Nishida *et al.* 2017, Numamoto *et al.* 2017). To our knowledge, the study by Horwitz *et al.* (2015) is the only published application of CRISPR-Cas9 in *K. lactis*. The authors report Cas9-mediated integration at three genomic loci simultaneously, which occurred at a rate of 2% in a NHEJ-deficient strain with an integrated Cas9 cassette using ca. 1000-bp homology flanks. However, data on targeting efficiency were not provided, preventing a meaningful comparison with the results obtained by the pUDP system in this study (96% targeting efficiency, 31% HR-rate, wild-type *K. lactis* strain, ca. 480-bp homology flanks, single locus). However, the absolute quantification of the HR-mediated repair left no doubt regarding pUDP tool efficacy for gene editing in *K. lactis*. In *K. marxianus* NBRC 1777, the pUDP system achieved higher targeting efficiencies (97% vs 65%) and similar efficiencies of marker-free HR (22% vs 38%) compared to the most efficient CRISPR system for this yeast reported to date (Nambu-Nishida *et al.* 2017). However, in this comparison it should be taken into account that, in this study, longer homology flanks were used (480 bp vs 50 bp), which may have facilitated HR-mediated repair. The published CRISPR-Cas9 tool for *O. polymorpha* (Numamoto *et al.* 2017) used strain BY4330, a mutant derived from the *O. polymorpha* NCYC495 background which is closely related to strain CBS 4732 employed in this study. Compared to this tool, the pUDP system achieved similar efficiencies of NHEJ-mediated gene disruption when compared with *O. parapolymorpha* CBS 11895 (61% vs 71%), but was less effective when compared with *O. polymorpha* CBS 4732 (9% vs 71%). In contrast to our findings, the authors were able to obtain gene-disrupted and marker-integrated transformants directly on transformation plates, indicating either a more suitable expression of gRNA and/or Cas9 or a different genetic tractability of the strain lineages. Compared to the pUDP system, all three discussed CRISPR-Cas9 tools (partially) rely on species-specific genetic parts. The pKD1 stability element was employed for plasmid replication in *K. lactis*

(Horwitz *et al.* 2015), while the *KmSNR52* promoter was used for gRNA expression in the case of *K. marxianus* (Nambu-Nishida *et al.* 2017), and an *OptRNA*^{Leu} promoter for gRNA expression in *O. polymorpha* (Numamoto *et al.* 2017). These RNA polymerase III-dependent promoters are unlikely to function in other species or genera.

RNA polymerase III-driven *Spcas9*-based editing strategies, such as those based on chimeric gRNA expression systems, are not easily compatible with multiplexed genome editing. Hitherto, the maximum number of gRNAs shown to enable simultaneous DSBs at different loci in *S. cerevisiae* is six. However, this result required an elaborate construction scheme using three plasmids for the individual expression of each gRNA (Mans *et al.* 2015). In contrast, ribozyme-flanked gRNA expression, as previously reported in *S. pastorianus* (Gorter de Vries *et al.* 2017), allowed to achieve the first double gene editing in *O. parapolymorpha* using a polycistronic gRNA array including two ribozyme-flanked spacers. The RNA polymerase II expression certainly facilitated to increase and to a greater extend modulate the expression of the gRNA. The expression of multiple gRNAs could be used as well to enhance gene deletion, even in absence of efficient HR, by providing the possibility to induce double cuts around the target gene which could result in a complete gene deletion even when repaired by NHEJ. Exploring and increasing the efficiency of multiplexed editing ability will be instrumental to further unlock genetic tractability of non-conventional yeasts.

In addition to the challenge of introducing DSBs quickly and efficiently, CRISPR-Cas9-mediated genome engineering in non-conventional yeasts is limited by their intrinsic ability for HR-mediated DNA repair, as illustrated in this study and by other previously developed CRISPR tools (Nambu-Nishida *et al.* 2017, Numamoto *et al.* 2017, Weninger *et al.* 2016). While low HR activity can often be compensated for by working in a NHEJ-deficient background, such strains may have undesirable properties, e.g. higher stress sensitivity (Nielsen *et al.* 2008, Takahashi *et al.* 2006) and reduction of cellular fitness (Snoek *et al.* 2009), which might only emerge after prolonged cultivation. For example, the commonly targeted *KU70/80* complex is involved in telomere maintenance, and disruption of the complex results in telomere shortening (Boulton and Jackson 1996), altered position of telomeric DNA in the nucleus (Laroche *et al.* 1998) and promotes telomere degradation and recombination (Polotnianka *et al.* 1998). As an alternative to inactivating NHEJ, Rad52, a highly conserved protein which plays an important role in HR-mediated DNA repair in yeasts and other organisms (Symington 2002), has been exploited to improve the efficiency of HR. Expression (Di Primio *et al.* 2005) or protein delivery (Kalvala *et al.* 2010) of *S. cerevisiae*-derived Rad52 as well as the utilization of a recently described Cas9-*Sc*Rad52 fusion protein (Shao *et al.* 2017) increased HR rates in mammalian cells. However, overexpression of *Kl*Rad52 in *K. lactis* did not have a beneficial effect (Kooistra *et al.* 2004). Another approach to optimize HR could be cell cycle synchronization and transformation of cells

in S-phase, during which cells exhibit the highest ratio of HR over NHEJ (Barlow and Rothstein 2010, Chapman *et al.* 2012, Karanam *et al.* 2012). Cell arrest in S-phase with hydroxyurea indeed led to increased rates of HR in various (non-conventional) yeasts (Tsakraklides *et al.* 2015).

To conclude, the pUDP CRISPR/Cas9 system presented here has been shown to work in four different species of yeasts belonging to the *Saccharomycotina*, limited by host-dependent targeting efficiency and intrinsic HR capability. We expect the pUDP system to be also applicable for Cas9-mediated genome engineering in other industrially relevant *Saccharomycotina* yeasts. The approach presented in this study demonstrates the potential of wide-host-range tools for genome editing. Although highly efficient genome editing is likely to require species-specific characteristics, the pUDP system can be used for rapid introduction of screenable mutations (e.g. auxotrophic markers) and, as demonstrated for *O. parapolymorpha*, the elimination of NHEJ. Moreover, due to its broad host range, the pUDP plasmid can serve as a starting point for optimization of Cas9-mediated gene targeting in individual yeast species by promoter and terminator replacement studies.

Acknowledgements

We thank Marcel van den Broek (Delft University of Technology) for the bioinformatics analysis and Philip de Groot (Delft University of Technology) for his contribution to the pUDP system.

Chapter 3

Evaluation of a novel cloud-based software platform for structured experiment design and linked data analytics

Hannes Jürgens*
Matthijs Niemeijer*
Laura D. Jennings-Antipov*
Robert Mans
Jack Morel
Antonius J. A. van Maris
Jack T. Pronk
Timothy S. Gardner

*Joint first authorship

Essentially as published in Scientific Data (2018); 5:180195
https://doi.org/10.1038/sdata.2018.195

Abstract

Open data in science requires precise definition of experimental procedures used in data generation, but traditional practices for sharing protocols and data cannot provide the required data contextualization. Here, we explore implementation, in an academic research setting, of a novel cloud-based software system designed to address this challenge. The software supports systematic definition of experimental procedures as visual processes, acquisition and analysis of primary data, and linking of data and procedures in machine-computable form. The software was tested on a set of quantitative microbial-physiology experiments. Though time-intensive, definition of experimental procedures in the software enabled much more precise, unambiguous definitions of experiments than conventional protocols. Once defined, processes were easily reusable and composable into more complex experimental flows. Automatic coupling of process definitions to experimental data enables immediate identification of correlations between procedural details, intended and unintended experimental perturbations, and experimental outcomes. Software-based experiment descriptions could ultimately replace terse and ambiguous 'Materials and Methods' sections in scientific journals, thus promoting reproducibility and reusability of published studies.

Introduction

Iterative progress in science and technology rests on the fundamental assumption that researchers can build off each other's published findings. Turning this assumption into reality can, however, be far from trivial for any field of science (Ioannidis 2005). Accurate descriptions of experimental procedures are essential for meaningful comparison, interpretation and re-use of data acquired in different laboratories, at different times, and/or by different researchers. However, 'Materials and Methods' sections in scientific journals are extremely terse, leaving room for subjective interpretation, and are often insufficient to accurately reproduce the stated findings. The magnitude of this challenge is illustrated by the frequent occurrence and estimated costs of irreproducibility in multi-laboratory studies in the life sciences (Baker 2016, Freedman *et al.* 2015, Gardner 2014), and the emergence of programmes that fund replication of impactful research, such as the Dutch Replication Studies pilot (NWO 2016).

Problems related to reproducibility and reusability of experimental data are further compounded by incomplete sharing of primary data and/or disconnection of data from the exact experimental context in which they were generated. For example, microbial biotechnology publications rarely include complete sets of primary data obtained in controlled-cultivation experiments (e.g., actual concentrations of metabolites in culture samples). Instead, publications generally only report derived parameters (e.g., biomass-specific conversion rates and yields) which might even be calculated differently by different researchers. Conversely, genome-wide 'omics' data derived from cultivation experiments are routinely deposited in publicly accessible repositories, yet the description of the exact experimental context in which data were generated is often limited and disconnected from the data. For example, transcriptome datasets generally only contain incomplete descriptions of experimental procedures, even though transcriptome analyses are notoriously sensitive to minor differences in experimental conditions, such as different sources of chemicals, cultivation procedures, and/or equipment settings (Baerends *et al.* 2008, Knijnenburg *et al.* 2009).

For a dataset to be truly reusable, all experimental data must be directly linked, in computable form, to the methodological information used in data generation. Data is considered to be in a "computable form" when each data point is annotated or linked to standardized data describing its experimental context and structured into standard formats suitable for data processing. Such data displays clear relationships between the variables and the observations (typically as rows and columns in a standard statistical data table) that allows search, quantitative filtering, categorization, summarization and correlation of data by visualization software (e.g., Spotfire, Tableau), statistical software (e.g., R, SAS, JMP) or automated data pipelines (built, for example, on Python or Apache

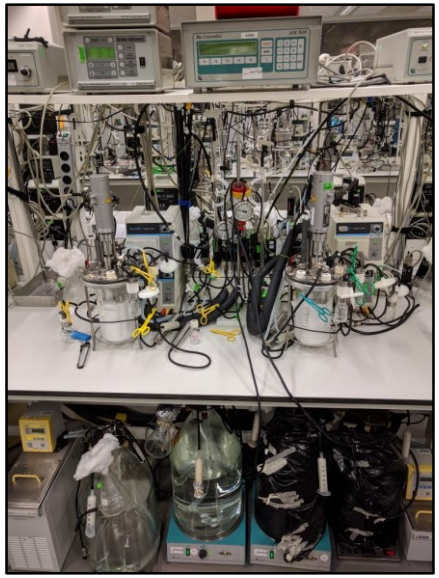
Spark). This format allows for meaningful reinterpretation and re-use of data through human or algorithmic analysis of data integrity and scientific validity. Computable, contextualized data sets also permit large-scale search and aggregation of data for use in increasingly powerful machine learning technologies.

Several scientific journals, including Scientific Data, are at the forefront of the drive towards open data publishing and data reusability (Wilkinson *et al.* 2016). Yet, these much-needed efforts are currently limited in their impact because tools that automatically link procedural information to measurement data in computable form are practically non-existent. General-purpose data repositories, e.g. FigShare (figshare.com), Dryad Digital Repository (datadryad.org), Mendeley Data (data.mendeley.com) and Zenodo (zenodo.org), facilitate access to data files via the cloud, but do not require computable methodological information to be published alongside datasets. Software like ISA Tools (Rocca-Serra *et al.* 2010) allow researchers to annotate datasets with metadata, thus increasing data discoverability. However, such procedures require extensive human annotation of data sets *after* data collection.

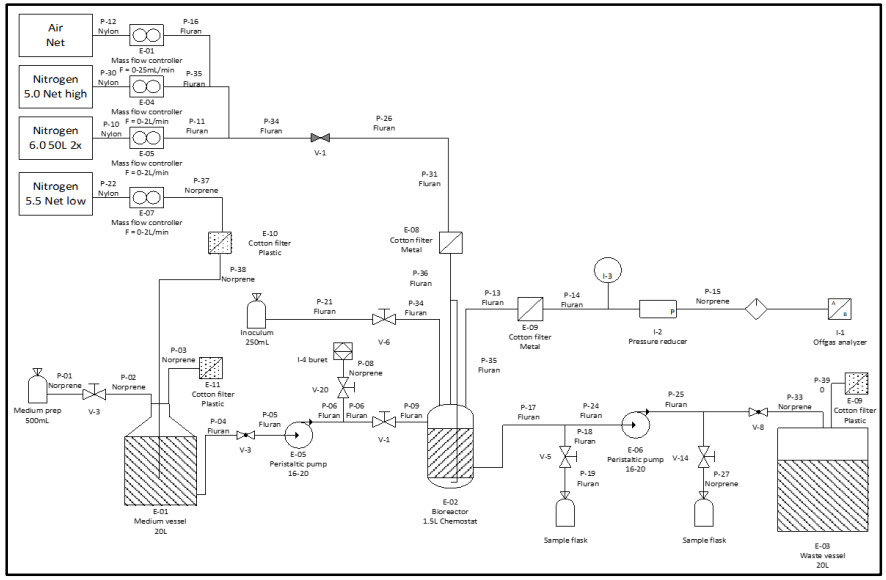
In 2016, Riffyn launched a new cloud-based software, the Riffyn Scientific Development Environment (SDE), for computer-aided design and analysis of experiments, with the goal of addressing challenges in data contextualization and reusability. Here, we evaluate this software in the practical setting of academic microbial biotechnology research, and discuss both its advantages and challenges for experiment design, data contextualization, and reuse of results in academic research. The evaluation was undertaken via a collaboration between scientists in the Industrial Microbiology section of the Delft University of Technology (TU Delft, the "evaluation team"), and scientists at Riffyn who support the Riffyn SDE application (the "support team"). The Riffyn SDE was used by the evaluation team, in their laboratory at TU Delft, to document, execute and analyze microbial cultivation in shake flasks, batch bioreactors and chemostat bioreactors.

Bioreactor cultivation offered a suitable use-case for the evaluation because it is a complicated, technically involved process (**Figure 3.1a-c**) that is presently documented with lengthy prose descriptions of the procedural steps and parameter settings involved. These documents are difficult to maintain and therefore frequently become out-of-date due to continuous evolution and improvement in cultivation methodologies. Moreover, even detailed protocols used within research groups can fail to fully capture experimental procedures, leading to confusion about how data were collected, and thus impairing future data analysis or quality assessment. The evaluation team, therefore, sought to assess if the Riffyn SDE would offer a meaningful improvement in the design, operation and analysis of such microbial biotechnology experiments.

a



b



c

Chemostat cultivation

Aerobic chemostat cultivation was performed in 2-l fermenters (Applikon, Schiedam, The Netherlands) at a working volume of 1 l and at 30°C. The ADI 1030 biocontroller kept the pH at 5. 0 by titration with 2 M KOH. A stirrer speed of 800 rpm and an air flow of 0. 5 l min^{-1} kept the dissolved oxygen tension higher than 40% of air saturation in all chemostat cultivations performed. The influx of medium was regulated by a peristaltic pump. The working volume of the cultures was kept constant by means of an electric level sensor. Cultures were assumed to be at steady state when, after at least five volume changes, the relevant parameters, such as culture dry weight, glucose concentration, ethanol concentration and respiratory quotient (i. e. the ratio between carbon dioxide production and oxygen consumption), changed by less than 2% during one volume change, and showed no downward or upward trend in the last three measurements, at least one volume change apart. Sustained oscillations of the dissolved oxygen concentration were not observed. There was no significant difference in dry weight (<1%) between effluent and culture liquid.

Figure 3.1: (a) Photograph of a bioreactor setup for chemostat cultivation, which combines equipment for aseptic liquid/gas handling and maintenance of a stable environment for microbial cultures, as well as measurement and data logging devices. **(b)** Flowsheet that might be used to communicate a bioreactor setup to other researchers, containing information about the type of mechanical bioreactor components used. **(c)** A typical text describing chemostat cultivation in a Materials and Methods section of a scientific publication, outlining the conditions and settings most important to the experiment.

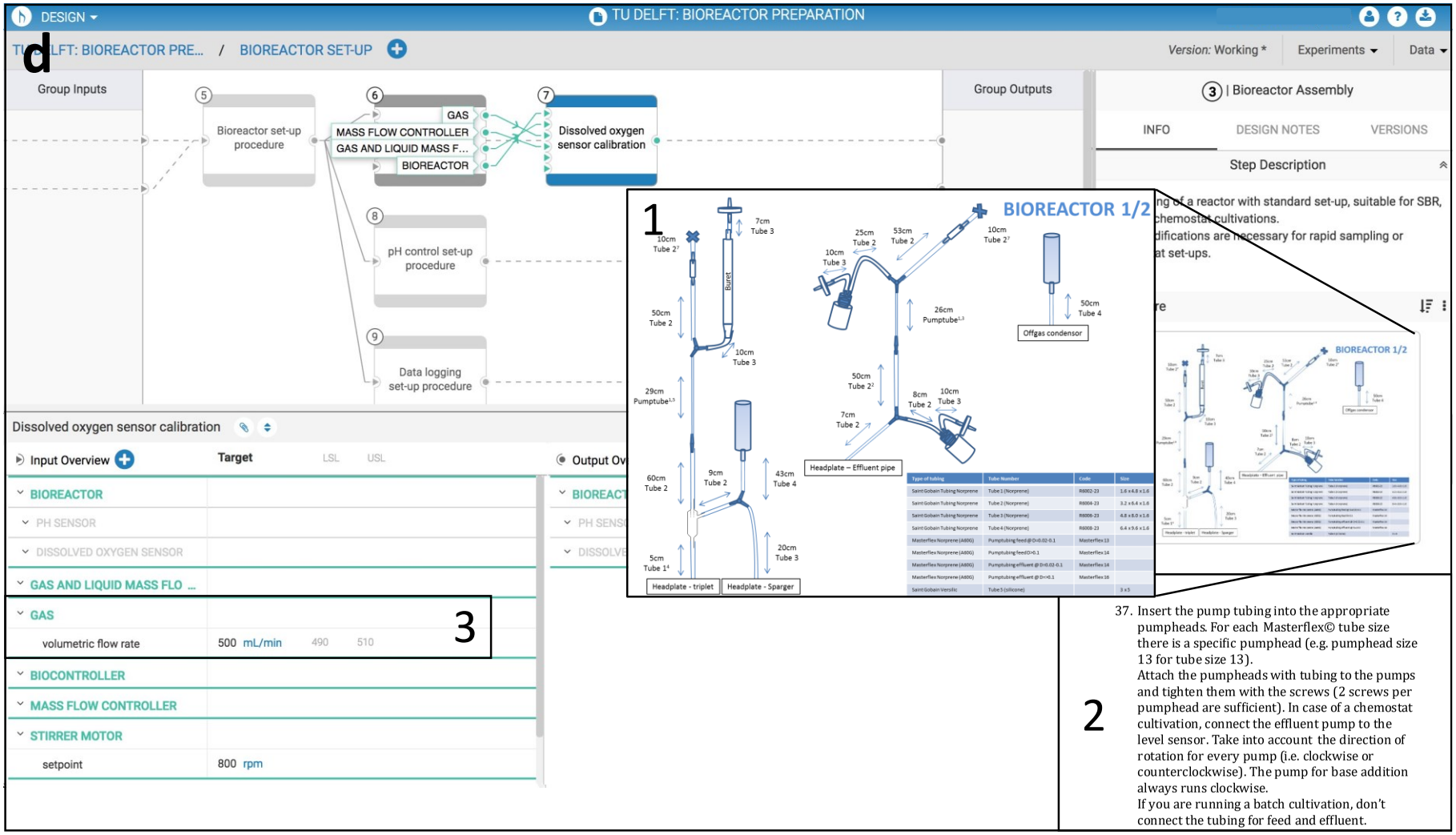


Figure 3.1, continued: (d) The Riffyn SDEs Design Mode allows the capture of illustrations to demonstrate complex bioreactor set-ups (d1), as well as detailed step-by-step instructions for such setups (d2), and target parameter values and their upper and lower spec limits for specific bioreactor settings (d3).

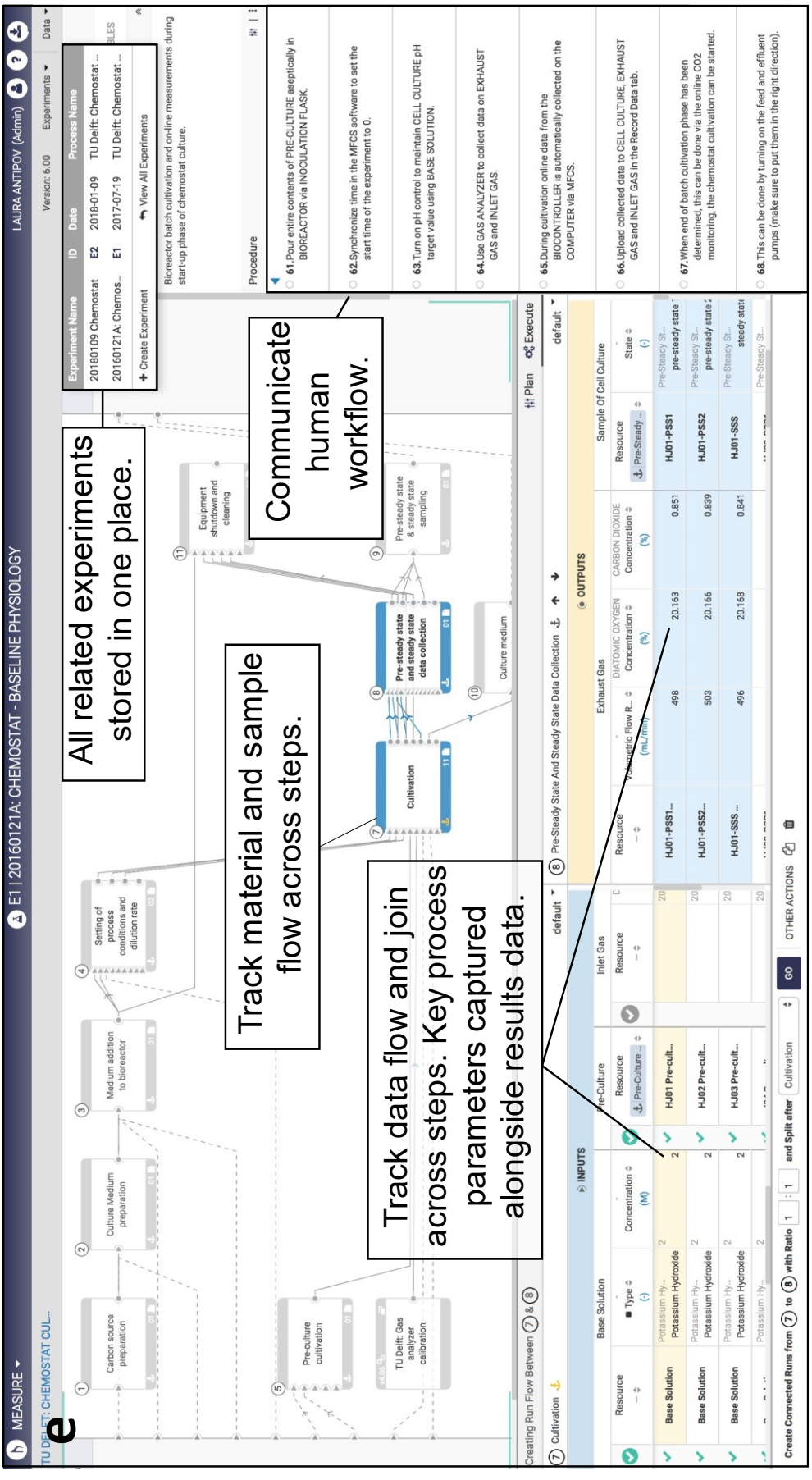


Figure 3.1, continued: (e) The Riffyn SDEs Measure Mode allows the association of data with process design backbones. Material, equipment and sample flow across steps are tracked; data associated with these entities are also tracked and joined across steps. Key process parameters, traditionally captured in lab notebooks, are captured electronically alongside results data otherwise captured in instrument files or spreadsheets. The human workflow needed to execute the experiment, traditionally recorded in lab notebooks or by word of mouth is recorded in context on each step. Also illustrated in this panel (upper right) are multiple experiments captured on this same process which will be collated together automatically during data analysis.

Riffyn Scientific Development Environment (SDE). The Riffyn SDE is a cloud-based software system for computer-aided design of experiments and data analysis. It is designed to enhance experimental data reproducibility and reusability by providing automatic data contextualization, and structured, computationally mineable, reusable datasets.

Data contextualization occurs in the Riffyn SDE by capturing data in three different flows: (1) material and equipment flow, (2) human workflow, and (3) data flow (**Figure 3.1d-e**). This data capture is achieved by transforming experimental protocols into reusable templates called "processes". Processes are created in the Riffyn SDE's "Design Mode", wherein experimental procedures are recorded as stepwise process flow diagrams, and the flow of materials and equipment across steps is systematically tracked. All relevant settings and key defining characteristics of these materials and equipment are specified directly in the process design. To communicate the human workflow needed to complete an experiment, detailed procedural instructions are also associated with each step of the process.

Each process can then be copied as one or more experiments, which are executed in the Riffyn SDE's "Measure Mode". Each of these copies reflects and preserves the version of the process at the time it was created, and becomes the computable backbone to which the experimental parameters, sample identifiers, and data are attached. Because experimental data and procedural parameters are recorded on the same backbone, measurement data (such as metabolite concentrations) get automatically and permanently joined to the protocol variables (such as HPLC mobile phase) used to obtain those data. Thus, experimental data are automatically contextualized, eliminating the need to annotate datasets with metadata after the experiment is complete.

After annotation with procedural information, the Riffyn SDE automatically joins the data across processing steps to provide the full provenance and traceable genealogy of materials, samples and reagents used as inputs to the experiments (**Figure 3.2**). The data joining performed by the Riffyn SDE is distinct from traditional data linking, which is often in the form of hyperlinks. In the Riffyn SDE, a join produces a standard statistical data frame ready for immediate visualization, statistical analysis and machine learning in common software such as Python, R and JMP. The Riffyn SDE joins data in three ways: (1) "horizontally" across all steps in a process flow diagram, (2) "vertically" across multiple experiments executed on a process (data concatenation), and (3) across multiple processes connected by shared sample outputs and inputs. The result is a comprehensive statistical data table containing all methodological and experimental data collected on a process or group of processes. This data table is searchable and computationally mineable, enabling immediate analyses of data quality, correlations, time-series analysis, causal relationships, and root causes of error.

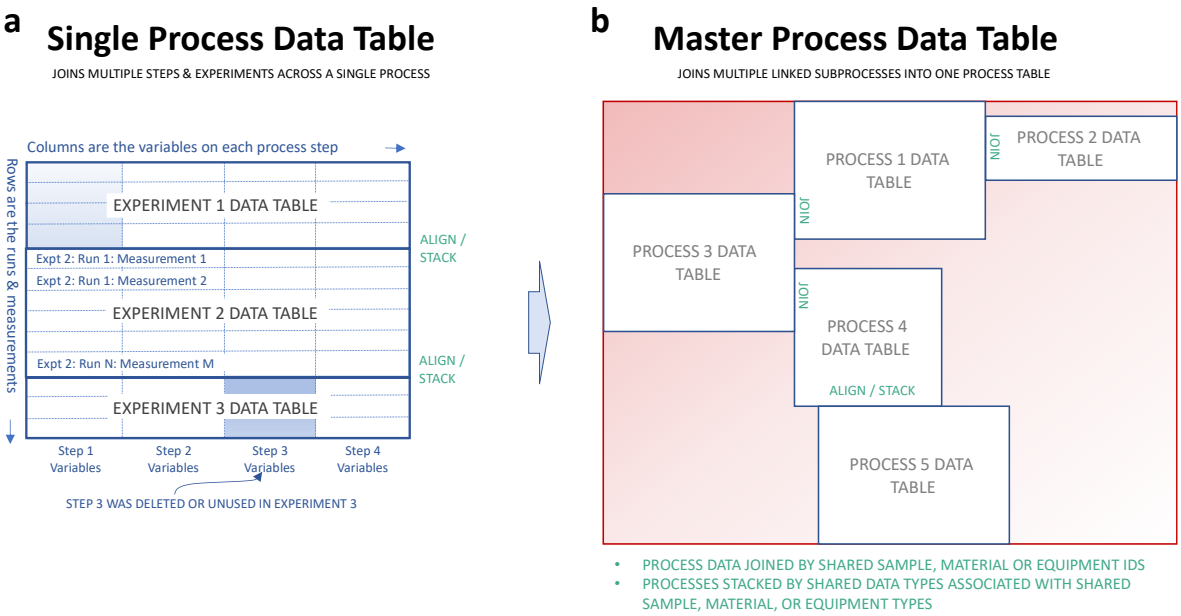


Figure 3.2: (a) All data on each experiment in the Riffyn SDE is extracted and flattened into a statistical data-frame compatible with nearly any modern analysis software (including R, JMP, SAS, Tableau, Minitab, etc.). This data-frame is composed of all variables defined on a Riffyn SDE process or experiment, including sample identifiers, parameter settings, start/stop times, dates, measurement data and units. The variables are presented as columns in the data-frame and grouped by step in the process. Multiple experiments executed on that same process (regardless of version) are stacked together such that data are aligned into common columns. If versions of a process diverge, gaps are left in the table where such variables are missing in the process. **(b)** Individual process data tables can be further combined into a "Master Process" data table composed of all data from all linked processes used in performing an experiment, including material preparation, equipment set-up, fermentation, and analytical chemistry processes. Data from these processes are joined together using the material flow graph defined by the Master Process in the Riffyn SDE, and the material identifiers automatically tracked by the Riffyn SDE.

Although bioreactor experiments were used as a test case in this study, the Riffyn SDE is not restricted to this domain of science. The Riffyn SDE has to-date been used for a variety of experiments including enzyme assays, HPLC assays, genetic engineering protocols, compound and material formulations studies, material preparation processes, cell line screening, animal pharmacokinetics, drug tablet manufacturing process development, food processing, and chemical or protein recovery and purification processes. Examples of some of these processes are provided as public processes on the Riffyn cloud at https://app.riffyn.com.

Riffyn SDE System architecture. The Riffyn SDE is a cloud-based software-as-a-service architecture hosted and operated by Riffyn in both multi-tenant and single-organization configurations on Amazon Web Services. Its graphical user interface is accessed manually via a web browser, or programmatically via a web API. The system is built on a foundation of NodeJS and Python language with the MeteorJS framework, SVG, D3 and WebGL graphics, OpenSSL encryption, and NginX load balancing. The Riffyn SDE software automatically saves all edits and data uploads, and concurrently updates all active web browsers such that users can see each other's changes in real-time. All

data remains encrypted throughout the system, both in transit and at rest using TLS 1.1 and 1.2 protocols over https and websockets with a variety of cipher keys. Data is stored in a MongoDB with three-fold real-time failover protection, a fourth real-time replica database, and static full-database backups every 6 hours, which are saved for 6 months to an offsite location for disaster protection. Data storage provides assured consistency on the primary member of the replica set, with eventual consistency to the secondary members of the replica set. The system automatically distributes load over a minimum of 18 servers with three-fold redundancy, and is horizontally scalable as load increases over time. Multimedia is stored in Riffyn as raw (original) files on AWS S3 buckets, ensuring scalability.

Data captured in the Riffyn SDE is automatically parsed, atomized and stored in a flexible No-SQL database as JSON document objects, using a hypergraph data model that links all data to the material and equipment sources that generated it (Gardner 2016). The Riffyn SDE integrates standard scientific ontologies from public sources such as BioPortal (Noy *et al.* 2009), ensuring standardization of terminology used to annotate experiments. The data model also allows for the capture of single-point data (e.g. microtiter per-well measurements), time series data (e.g. batch bioreactor data) and event series data (e.g. chromatography spectra and flow cytometer data) in the same underlying data structures. When data is requested by the user, the Riffyn SDE analytics engine walks the underlying material hypergraph, automatically annotates measurements with experimental metadata, links them based on source/child material genealogy, and reshapes them into the flat data tables suited for immediate visualization and statistical analysis.

Data exported from the Riffyn SDE is provided in vendor-neutral CSV, JSON or XML format and provides a complete description of both experimental method and measurement data, such that users are not bound by proprietary data formats or storage. All data can be streamed in real-time to external data stores, maintained independently of the Riffyn hosting service. A schematic re presentation of this data capture, archival and joining is show in **Figure 3.3**.

Riffyn is capable of handling large amounts of data on each experiment and each sample collected. Prior examples include writing 10,000 measurements to a single sample in some long-running temporal experiments, and collecting and tracking 10,000 samples in a screening experiment. Riffyn is currently extending its capabilities to permit tracking of 100,000 samples per experiment (by autumn 2018), but high-throughput screening experiments of a larger scale (millions of samples per experiment) require more specialized automation software.

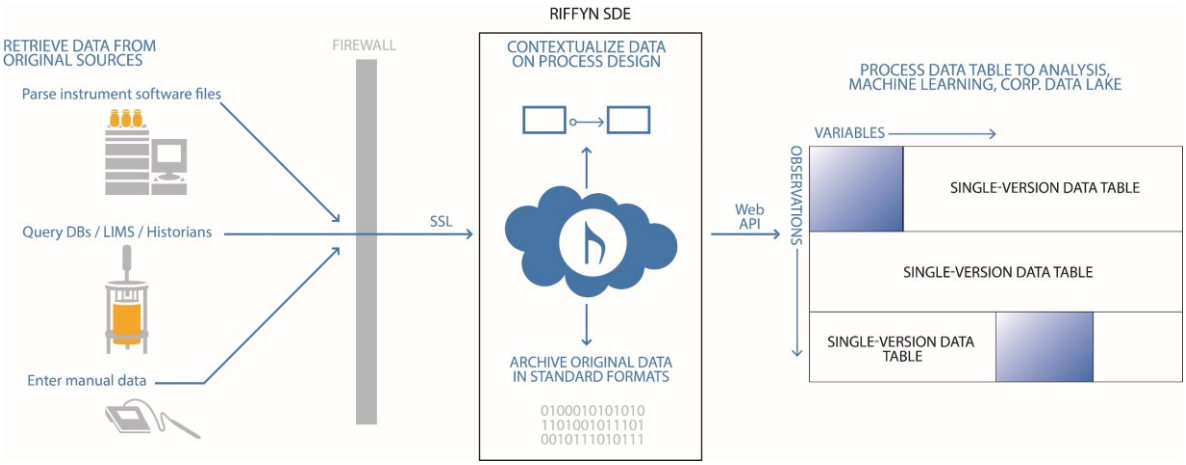


Figure 3.3: Data is entered into Riffyn via parsing of instrument files, querying of a database, or via manual data entry. All data is passed through a firewall and contextualized based on the Riffyn SDE process design. Any files parsed during the data upload are additionally stored in their native format. All contextualized data is exported as a flat csv file with all data, material information, and setpoints captured. This flat file is built with each variable as a column, and each row as an observation. Multiple tables are concatenated together across each version of the process to provide a comprehensive, cross-version table that captures all data across the development of the process.

Experiments performed in this study. To evaluate the ability of the Riffyn SDE software to capture and communicate complex protocols, record diverse data types, and enable data analysis, sharing, and reusability, we applied it to a study to quantify the aerobic growth characteristics of the industrially relevant yeast *Ogataea polymorpha* (syn. *Hansenula polymorpha*). This methylotrophic and thermotolerant yeast is used at an industrial scale for biotechnological production of recombinant proteins (Gellissen 2000), and has several interesting properties for application in other processes. These properties include an ability to grow at temperatures up to 50°C, consumption of a wide range of substrates, low byproduct formation in aerobic cultures and a generally recognized as safe (GRAS) status (Kunze *et al.* 2009, Kurtzman 2011). Despite its industrial applications, and many academic studies of its growth on methanol (van Dijken *et al.* 1976, van Zutphen *et al.* 2010) and nitrate (Avila *et al.* 1995, Pignocchi *et al.* 1998), aerobic growth characteristics of *O. polymorpha* on glucose have not been described comprehensively. Here, the aerobic physiology of two popular *O. (para)polymorpha* strains (CBS4732 and CBS11895/DL-1) was investigated in a glucose-containing synthetic medium (Verduyn *et al.* 1992). In a first set of experiments, shake-flask cultures were used to determine specific growth rates over a range of temperatures. Subsequently, a precise quantitative analysis of growth kinetics was performed in both bioreactor batch cultures and chemostat cultures.

In shake-flask cultures, cells were added to sterile synthetic growth medium and incubated aerobically at temperatures ranging from 30 to 49°C. Optical density measurements were performed at regular intervals to enable calculation of the maximum specific growth rate at each temperature. Subsequently, both strains were characterized in bioreactor batch cultures at 30°C (often used for characterization of

yeasts) and 40°C (optimum growth temperature determined in shake flasks). Bioreactor batch cultivation allowed for control of medium pH, continuous off-gas measurement, as well as intensive mixing and gas transfer to prevent oxygen limitation of fast-growing microbial cultures. Samples from the bioreactor cultures were used to determine biomass concentration and analyzed by HPLC to quantify the extracellular metabolite concentration. Finally, both strains were grown in chemostat cultures. In a chemostat, simultaneous addition of fresh medium to the bioreactor and removal of spent culture broth occurs at a fixed dilution rate D [h^{-1}], defined as the outgoing liquid flow F_{out} [L h^{-1}] divided by the liquid culture volume V_L [L] in the reactor ($D = F_{out}/V_L$ [h^{-1}]). When D does not exceed the highest specific growth rate μ [h^{-1}] that the organism can reach under the experimental conditions, a nutrient-limited steady state will ensue in which μ equals the dilution rate set by the experimenter (Hakkaart *et al.* 2017). In this study, physiology of both *O. (para)polymorpha* strains was compared at 30 and 40°C in steady-state chemostat cultures operated at D = 0.10 h^{-1}.

Results

Experiment design in Riffyn SDE Design Mode. The Riffyn SDE's Design Mode was evaluated by applying it to microbial cultivation in shake flasks and bioreactors, the latter in batch as well as chemostat mode.

Design Mode was used to define a process comprising detailed, step-by-step procedural instructions, visualized by process flow diagrams in which all required materials and equipment are specified along with relevant settings and other process parameters. For example, the process for bioreactor cultivation includes detailed descriptions of the composition and preparation of synthetic growth media and of pH setpoints and control during cultivation. Moreover, steps that are difficult to explain in written protocols, can be clearly defined by embedding videos or photographs into the Riffyn SDE process diagram. For instance, the procedure for linking two stainless-steel connectors during coupling of a new medium reservoir to a bioreactor, which is difficult to describe in words but essential for aseptic operation of chemostat cultures, was adequately described by a short video embedded on the corresponding step 3 of the Process (https://app.riffyn.com/processes/24K9JeAQg6vJHPrg5, see Methods section for access to Riffyn SDE).

Although definition of a process involved significant time investments, this activity encouraged the evaluation team to critically assess and discuss all individual steps in an experimental procedure and describe them unambiguously. For example, definition of a process for chemostat cultivation inspired more rigorous definition and/or standardization of the length and diameter/type of tubing used on bioreactors, the procedure for setting pump-speeds for chemostat setups, the dead volume of sampling ports, and the treatment of off-gas before analysis. These insights were put to good use when, in the midst of the software evaluation, a physical move of the laboratory to a new building necessitated the complete disassembly and reassembly of all its 40 bioreactor set-ups. At a deeper level, the design phase, which precedes and defines the actual execution of the experiments, compelled researchers to think critically about which experimental parameters to collect and record. This intrinsic link between experimental design and execution allows users to improve interpretability of experiments. For example, the procedures in the inoculation phase, an essential step in microbial cultivation that involves experimental parameters such as volume, biomass concentration, physiological status and means of addition of a preculture, can drastically affect experimental outcome but are rarely reported in literature in a completely unambiguous manner.

The Riffyn SDE software automatically tracks changes to protocols and allows generation of a new version at any time. This ensures that the most up-to-date and/or most suitable variant of a process is available to experimenters, thereby reducing the

amount of potential errors and mistakes. Change tracking also ensures that (subtle) changes in an experimental process can be correlated retrospectively to changes in experimental outcome.

Data collection in Riffyn SDE Measure Mode. The processes defined in Design Mode were used as templates for data collection in Measure Mode to characterize aerobic growth of *O. polymorpha* on glucose in shake flasks and in bioreactor batch and chemostat cultures. In the evaluation team's laboratory, data are usually recorded in laboratory notebooks and later combined in Excel for calculation of key experimental outcomes. By contrast, in Measure Mode, the experimenter is systematically guided through each step of the experiment and prompted to collect parameter values (defined in Design Mode), such as equipment settings, amounts of reagents used and associated experimental measurement data (e.g. on-line off-gas measurements and metabolite/biomass concentrations) in real time.

In contrast to the classical use of lab journals, where data are usually recorded on a 'by date/time' basis, the Riffyn SDE's Measure Mode stores all measured data for a particular experiment together. This is especially relevant and helpful when cultures are tracked over time (e.g. batch cultivation experiments that cover multiple days), where the primary recorded data often become fragmented when multiple experiments are performed in parallel. Via a built-in plotting tool, newly obtained measurements could be directly integrated with data obtained at earlier time-points. This feature could be very useful for identifying sampling or analysis errors on the spot, which could prompt the experimenter to take another sample.

A useful feature of Measure Mode was the option to make changes to process designs while performing an experiment. This possibility is especially valuable when an experimenter is confronted with an unexpected development, such as a sudden change of supplier of a chemical or the need to freeze and store samples because analytical equipment is malfunctioning. Since the Riffyn SDE software unambiguously links any such modifications in the process design to the relevant data, their potential impacts on experimental outcomes remain unambiguously documented and can be retrospectively identified.

Data analysis. In the evaluation team's laboratory, students and employees typically manually extract and combine relevant data generated in their experiments, often into individually prepared spreadsheets, to perform calculations and compare experimental outcomes. This approach requires constant checks for consistency and accuracy, which can be inconvenient and time-consuming when applied to large experiments and datasets.

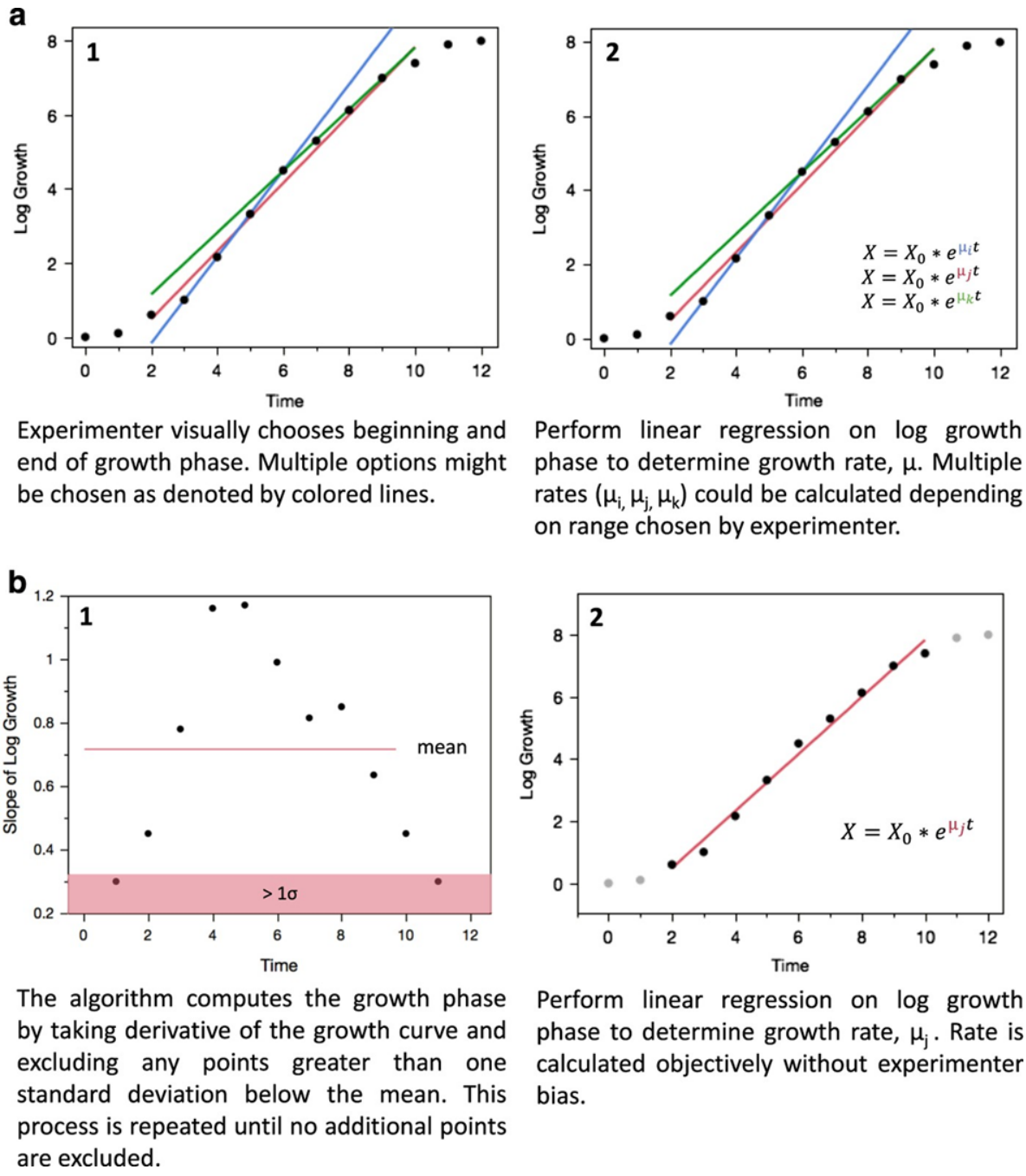


Figure 3.4: (a) The ad hoc manual method typically used by experimenters requires visual inspection of the growth curves to exclude data points believed to be outside of exponential phase. After exclusion of such points, growth rates are calculated by regression of the remaining data points. This approach is subject to subjective interpretation and human bias. **(b)** The automatic growth rate calculation method implemented in scripts associated with the cultivation processes in the Riffyn SDE. An algorithm, rather than visual inspection, is used to identify values outside the exponential growth phase. Growth rates are then calculated using the remaining data points. Physiological parameters derived from such growth rates are objectively calculated, without human bias.

The limitations of this approach become especially evident when checking and/or integrating data obtained at different times, by different researchers, and/or in different laboratories. Recording process parameters and experimental data with the Riffyn SDE Design and Measure Modes streamlined and standardized data capture and analysis. The software automatically generated CSV-formatted data tables for each process, combining not only the process parameters and experimental data for each step, but also from other experiments that were conducted using a specific process (**Figure 3.2**).

Even when starting from the same dataset, different researchers may adopt different ways to calculate derived parameters and perform statistical analyses. This problem especially occurs when calculations are based on a manually selected subset of the experimental data. For example, calculations of specific growth rates in microbial batch cultures are based on biomass and/or optical density measurements taken during the "exponential growth phase", with the exponential growth phase being defined in an ad hoc way by visual inspection of data by each researcher. The Riffyn SDE software resolved this issue of human bias by enabling an objective, automated data calculation pipeline for all physiological parameters (**Figure 3.4**). This pipeline utilized the automatically-generated Riffyn data tables in combination with a JMP software script, both of which are archived together with the experiment in Measure Mode. For example, the specific growth rate calculation script automatically defined the exponential phase of growth curves based on objective criteria, and then computed specific growth rates from those phases. This automated analysis was used to calculate specific growth rates, biomass yields and biomass-specific glucose uptake rates of two *O. (para)polymorpha* strains (CBS4732 and CBS11895) in aerobic bioreactor batch cultures at 30 and 40°C (**Table 3.1**) as well as for their quantitative physiological analysis in shake-flask and chemostat cultures (**Supplementary tables S3.1 and S3.2**). A summary of the design of the analysis scripts, and links to the full scripts within the Riffyn SDE, are provided in the Methods section.

Table 3.1: Temperature dependent physiology of two wild type *Ogataea (para)polymorpha* strains in aerobic bioreactor batch cultures.

	30°C		40°C	
	CBS4732	**CBS11895**	**CBS4732**	**CBS11895**
$\mu_{max,\ dryweight}$ [h^{-1}]	0.45 ± 0.01	0.36 ± 0.01	0.68 ± 0.01	0.63 ± 0.01
	± 0.008	*± 0.006*	*± 0.012*	*± 0.010*
$\mu_{max,\ OD660}$ [h^{-1}]	0.44 ± 0.01	0.36 ± 0.01	0.69 ± 0.01	0.63 ± 0.01
	± 0.001	*± 0.001*	*± 0.001*	*± 0.001*
$Y_{X/S}$ [g biomass (g glucose)$^{-1}$]	0.51 ± 0.00	0.52 ± 0.00	0.49 ± 0.01	0.49 ± 0.01
	± 0.01	*± 0.01*	*± 0.01*	*± 0.01*
$q_{Glucose}$ [mmol (g biomass)$^{-1}$ h^{-1}]	-4.90 ± 0.02	-3.90 ± 0.09	-7.77 ± 0.30	-7.15 ± 0.06
	± 0.15	*± 0.15*	*± 0.15*	*± 0.15*

Reported values are Mean ± MAD. Standard errors are reported in italics on the second line. Cultures were grown at pH 5 in synthetic medium with an initial glucose concentration of 7.5 g L^{-1}. Means and mean absolute deviations (MAD) were calculated from two individual cultures and the standard error was calculated from the pooled estimate of variance across all culture samples from each strain. For some properties subject to multiplicative error (e.g., growth rate), logarithmic transformations of the data were applied to regularize the variance prior to pooling. Data was transformed back to standard linear scale prior for reporting the errors. Symbols: $\mu_{max,\ dryweight}$ = maximum specific cell growth rate based on measurement of biomass dryweight; $\mu_{max,\ OD660}$ = maximum specific cell growth rate based on measurement of culture optical density at 660 nm, $Y_{X/S}$ = yield of cell biomass dryweight on sugar carbon source in the exponential phase, $q_{Glucose}$ = biomass-specific glucose uptake rate in the exponential growth phase.

Discussion

The Riffyn SDE offers a novel process-based approach to scientific experimentation, with the ability to integrate method definitions, primary research data, and data analysis scripts into a shareable representation of experiments. These visually expressive process descriptions increase the re-workability of complex procedures and the reusability of the data they generate. They are also computable which facilitates their use by search, data mining and data analysis programs. This approach could potentially address an Achilles heel of open data policies in academic research and publishing: the exchange of poorly contextualized, subjectively analyzed, and therefore intrinsically non-reusable data.

This evaluation illuminated some challenges in the initial implementation of the Riffyn SDE into a real-life academic research setting, mostly during first definition of experimental procedures. Firstly, getting acquainted with the software is a substantial up-front time investment for new users. Secondly, design of experiments in Riffyn is more time intensive than writing conventional protocols in prose format, as the designs are more complex and provide the framework for subsequent experimental execution and data analysis within the software. Furthermore, this connection between design, data collection and analysis can lead to the need for multiple design cycles for design optimization, unless the user is very experienced. However, once finalized, the procedures defined in Riffyn provide a definition of the experimental procedures that is more complete and unambiguous than traditional protocols. Therefore, given our experience during the execution and communication of this trial experiment, we speculate that software like the Riffyn SDE could helpfully augment or replace descriptions in methods sections of scientific publications.

One particular unanticipated benefit experienced by the evaluation team related to experimenter awareness. Implementation of intensively-used procedures in the systematic process format led to identification of missing information in established protocols, which invited the experimenters to rethink and discuss experimental design, process descriptions and standardization. Outcomes of these discussions were then memorialized into a shared representation of the process.

The Riffyn SDE also helped to address the challenges of tracking process changes via its version control system. This provides assurance of stable representation of past experimental methods and their resulting data for future analysis. It prevents undesirable situations where the exact experimental procedures have to be reconstructed retrospectively from the notebooks and memories of different researchers. For example, its potential relevance is illustrated by the severe loss of time that the evaluation team incurred several years ago when their regular supplier of antifoaming agent ceased production. Subsequently, over a substantial period of time,

replacement antifoaming compounds were found to affect yeast physiology in a context-dependent manner. Had all these experiments been recorded with the Riffyn SDE software, it would have been simple to correlate, for example, biomass yields or growth rates to the type of antifoam used across a large number of experiments.

We also identified a clear trade-off between the level of process detail implemented and the required investments in scientist time. While every aspect of a laboratory procedure that can influence experimental outcome could theoretically be defined in a Riffyn SDE process, it may not be practical from a time and effort perspective to do so. Especially in sparsely staffed academic research settings, we foresee the largest return on investment of effort when applying the Riffyn SDE to processes that are technically involved, generate quantitative results, are intensively used within a group, and whose basic technical methodologies are not subject to frequent change. In such situations, the processes can be reused and modified for new experiments, resulting in time savings and higher-quality data. Bioreactor cultivation of microorganisms and its associated analyses, which were used as a model for this study, provided an excellent fit with this description. On the other hand, the volatile world of genetic engineering may well represent the opposite end of a scale. When new methods and protocols appear on an almost weekly basis, and the quality of outputs can be verified by genome sequencing or other diagnostic tools, rigorous implementation of protocols as Riffyn SDE processes may require too large an investment in time.

In the evaluation team's laboratory, written protocols for key methods are currently maintained, updated and approved by academic and technical staff members. Keeping these protocols up-to-date, as well as ensuring that the right versions are used by a continually changing population of PhD, MSc and BSc students requires intensive communication and traffic of document versions. Since the Riffyn SDE is cloud-based, updates instantly, and is not limited to a single user at a time, multiple researchers were able to simultaneously and interactively work on the same process or experiment. These real-time sharing capabilities of the Riffyn SDE provide mechanisms to streamline communication, revision and approval of experimental process definitions. Additionally, the Riffyn SDE's ability to integrate multimedia approaches in process descriptions could provide a valuable tool for training of new students and employees, thereby minimizing loss of experiments and improving laboratory safety.

Finally, we note that the Riffyn SDE encouraged more automated, reproducible, and objective analysis of experimental data produced in the study. By automatically extracting, combining and reshaping both methodological and measurement data for computational analysis, we were able to develop and apply computational scripts directly to the processes and experiments in the Riffyn SDE. Moreover, by compiling data across multiple experiments, the Riffyn SDE facilitates learning from historical data records. We feel that this approach could enable more rigorous outcomes for the evaluation team's laboratory and ensure more consistent, traceable, and defensible outcomes when shared with the broader research community.

Methods

Riffyn SDE experiments serve as the Methods section of this article. Riffyn supports open science and has made all data, methods, code and supporting information available as files in the Supplementary Material (Data citation 1) in both human readable and computable form. In addition, all data are freely available for viewing or downloading via the Riffyn SDE user interface (https://app.riffyn.com) using a web browser, and programmatically accessible via an open application program interface (API) using any software supporting secure http requests (api.riffyn.com). Readers may request (via https://riffyn.com) an open access online account in the Riffyn SDE to view these experiments and create their own processes and experiments[4]. Riffyn provides a copy of its open access policy at https://riffyn.com/riffyn-open-access-policy.

Riffyn SDE processes. Processes used in this study, including material preparation, cultivation and analytical methods, were composed into an aggregate set of linked processes supporting each type of cell culture modality in this study (shake flask, batch bioreactor and chemostat cultivation). Each set of linked processes is called a "Master Process" (see **Figure 3.2**). Links to each of these Master Processes, and the experiments performed on the sub-processes within them, are provided below for reference. These experiments can also be found by browsing the Master Process structure itself via the Riffyn SDE.

- Shake flask batch cultivation Master Process:
  https://app.riffyn.com/processes/FonopPptQjMqQMBWm
    - Shake flask batch cultivation experiment:
      https://app.riffyn.com/experiments/ABpJX7NjeQJy9xBTK
- Bioreactor batch cultivation Master Process:
  https://app.riffyn.com/processes/czj5FuAfazh6if8j2
    - Medium preparation experiment:
      https://app.riffyn.com/experiments/EDuGj2GkQMke6vuEw
    - Bioreactor preparation experiment:
      https://app.riffyn.com/experiments/aqHPvgSFm2apMtGMW
    - Bioreactor batch cultivation experiment:
      https://app.riffyn.com/experiments/BTpb9qytcd5qwPTau
    - Biomass dryweight analysis experiment:
      https://app.riffyn.com/experiments/AnxhXchsrjwpB2zWs

[4] As of 2022, app.riffyn.com is not available, and therefore cannot serve as Methods section of this article. As mentioned above, all data, methods, code and supporting information are available in the Supplementary material online (Data citation 1).

    - Metabolite analysis (HPLC) experiment: https://app.riffyn.com/experiments/irhLrZ3s4oCZT7NyC
    - Optical density measurement experiment: https://app.riffyn.com/experiments/h6XWEFyNSBmJXrrfS
- Chemostat cultivation Master Process: https://app.riffyn.com/processes/3YstPHztm3kH9nuwX
    - Medium preparation experiment: https://app.riffyn.com/experiments/xNojK3JS8LaEdp7aZ
    - Bioreactor preparation experiment: https://app.riffyn.com/experiments/ZGXT3PhyvXqFW8bg3
    - Chemostat cultivation experiment: https://app.riffyn.com/experiments/GHNJvwkrzeiSdbqme
    - Rapid sampling for extracellular metabolites experiment: https://app.riffyn.com/experiments/FntsvnGc838TTepJ7
    - Biomass dryweight analysis experiment: https://app.riffyn.com/experiments/qLqiBdFddeZqSE5aC

Analysis scripts. Three data analysis scripts used in this study are summarized here. They were each designed to analyze a "Master Process" data table which is a combined data table built by joining all data on a Master Process (see **Figure 3.2**). The Master Process data tables were created using the 'Get and join experiment data' function in a JMP Add-In for the Riffyn SDE (JMP Add-In available from https://help.riffyn.com). Each analysis script is stored on the Master Process on the "analysis" step.

Script 1: Calculation of results from shake flask batch cultivation data. This script performs a set of successive actions to clean and organize the data before calculating the growth rate of yeast cells based on optical density measurements. The sections of the script, denoted by the line number, are as follows: (1- 67) loading, sorting, and cleanup of data; (68-173) calculation of growth parameters and rates; (174-403) summarization and tabulation of results.

Script 2: Calculation of results from batch bioreactor cultivation data. This script performs a set of successive actions to clean and organize the data before calculating basic physiological parameters. The sections of the script, denoted by the line number, are as follows: (1-75) loading, sorting, and cleanup of data; (76-328) calculation of growth parameters and rates; (329-681) summarization and tabulation of results.

Script 3: Calculation of results from chemostat cultivation data. This script performs a set of successive actions to clean and organize the data before calculating basic physiological parameters. The sections of the script, denoted by the line number, are as follows: (1-97) loading, sorting, and cleanup of data; (98-462) calculation of growth parameters and rates; (463-711) summarization and tabulation of results.

Code availability. The analysis scripts described above are available as .jsl files ('JMP scripts used for automatic analysis of Riffyn-generated data files', Data citation 1). The scripts were written for JMP version 13.1 and can also be opened by any plain text editor.

Data citations

1. Juergens, H. *et al. figshare* https://doi.org/10.6084/m9.figshare.c.4001472 (2018).

Acknowledgements

We thank Matthew Cockerill for his help in conceptualizing the study and approach, and review of the manuscript, Erik de Hulster for his advice regarding bioreactor operation, Wijb Dekker for his contributions to Figure 1, Veronica Gast for conducting part of the experimental work described in this study, Eric Kolaczyk for his guidance on statistical methodology, and the Riffyn SDE engineering team for building and operating the software.

Supplementary material

Table S3.1: Temperature dependent physiology of two wildtype *Ogataea (para)polymorpha* strains grown in synthetic medium at pH 5 in aerobic glucose-limited chemostat cultures. Means and mean absolute deviations (MAD) were calculated from two individual cultures and the standard error was calculated from the pooled estimate of variance across all culture samples from each strain. Carbon recovery calculations are based on a biomass carbon content of 48% (w/w). BDL: below detection limit (0.01 mM).

	30°C		40°C		Pooled Std Error
	CBS4732	**CBS11895**	**CBS4732**	**CBS11895**	
D (h^{-1})	0.10 ± 0.00	0.10 ± 0.00	0.10 ± 0.00	0.10 ± 0.00	*± 0.00*
Reservoir glucose (g L^{-1})	7.47 ± 0.02	7.49 ± 0.01	7.42 ± 0.02	7.43 ± 0.02	*± 0.01*
Residual glucose (mM)	BDL	BDL	BDL	0.02 ± 0.00	*± 0.00*
$Y_{X/S}$ (g biomass [g glucose]$^{-1}$)	0.49 ± 0.00	0.51 ± 0.00	0.46 ± 0.00	0.46 ± 0.01	*± 0.01*
$Y_{X/O2}$ (g biomass [g O_2]$^{-1}$)	1.16 ± 0.02	1.35 ± 0.05	0.97 ± 0.00	0.99 ± 0.02	*± 0.02*
RQ	1.03 ± 0.00	1.05 ± 0.01	1.01 ± 0.00	1.04 ± 0.00	*± 0.00*
$q_{Glucose}$ (mmol [g biomass]$^{-1}$ h^{-1})	-1.16 ± 0.02	-1.08 ± 0.03	-1.20 ± 0.01	-1.21 ± 0.01	*± 0.02*
q_{CO2} (mmol [g biomass]$^{-1}$ h^{-1})	2.78 ± 0.06	2.44 ± 0.07	3.27 ± 0.02	3.26 ± 0.02	*± 0.04*
q_{O2} (mmol [g biomass]$^{-1}$ h^{-1})	-2.69 ± 0.05	-2.32 ± 0.08	-3.23 ± 0.02	-3.15 ± 0.02	*± 0.04*
C_X (g biomass L^{-1})	3.63 ± 0.02	3.84 ± 0.08	3.39 ± 0.04	3.41 ± 0.04	*± 0.05*
Carbon recovery (%)	98.2 ± 0.3	99.3 ± 1.2	100.5 ± 1.0	99.8 ± 1.0	*± 0.8*

Reported values are Mean ± MAD. Standard errors are reported in italics in the right column.

Table S3.2: Temperature dependent maximum specific growth rates (μ_{max}) of two wild-type *Ogataea (para)polymorpha* strains in aerobic shake flask batch cultures. Cultivation was done in synthetic medium with an initial glucose concentration of 20 g L^{-1}. Means and mean absolute deviations (MAD) were calculated from two individual cultures and the standard error was calculated from the pooled estimate of variance across all culture samples from each strain. Logarithmic transformations of the data were applied to regularize the variance prior to pooling. Values for CBS4732 at 49°C and CBS11895 at 48 and 49°C were derived from cultures directly inoculated from glycerol stocks.

	***O. polymorpha* CBS4732**					
	30°C	**37°C**	**40°C**	**45°C**	**48°C**	**49°C**
μ_{max}, OD_{660} [h^{-1}]	0.34 ± 0.00	0.60 ± 0.01	0.66 ± 0.00	0.61 ± 0.01	0.37 ± 0.02	0.20 ± 0.03
	±0.02	*±0.03*	*±0.05*	*±0.03*	*±0.02*	*±0.01*
	***O. parapolymorpha* CBS11895**					
	30°C	**37°C**	**40°C**	**45°C**	**48°C**	**49°C**
μ_{max}, OD_{660} [h^{-1}]	0.40 ± 0.00	0.51 ± 0.00	0.59 ± 0.00	0.52 ± 0.01	0.28 ± 0.00	0.18 ± 0.00
	±0.03	*±0.03*	*±0.04*	*±0.03*	*±0.02*	*±0.01*

Reported values are Mean ± MAD. Standard errors are reported in italics on the second line.

Chapter 4

Contribution of Complex I NADH dehydrogenase to respiratory energy coupling in glucose-grown cultures of *Ogataea parapolymorpha*

Hannes Jürgens*
Xavier D. H. Hakkaart*
Jildau E. Bras
André Vente
Liang Wu
Kirsten R. Benjamin
Jack T. Pronk
Pascale Daran-Lapujade
Robert Mans

*Joint first authorship

Essentially as published in Applied and Environmental Microbiology (2020); 86:e00678-20
https://doi.org/10.1128/AEM.00678-20

Abstract

The thermotolerant yeast *Ogataea parapolymorpha* (formerly *Hansenula polymorpha*) is an industrially relevant production host that exhibits a fully respiratory sugar metabolism in aerobic batch cultures. NADH-derived electrons can enter its mitochondrial respiratory chain either via a proton-translocating Complex I NADH dehydrogenase or via three putative alternative NADH dehydrogenases. This respiratory entry point affects the amount of ATP produced per NADH/O_2 consumed and therefore impacts the maximum yield of biomass and/or cellular products from a given amount of substrate. To investigate the physiological importance of Complex I, a wild-type *O. parapolymorpha* strain and a congenic Complex I-deficient mutant were grown on glucose in aerobic batch, chemostat, and retentostat cultures in bioreactors. In batch cultures, the two strains exhibited a fully respiratory metabolism and showed the same growth rates and biomass yields, indicating that, under these conditions, the contribution of NADH oxidation via Complex I was negligible. Both strains also exhibited a respiratory metabolism in glucose-limited chemostat cultures, but the Complex I-deficient mutant showed considerably reduced biomass yields on substrate and oxygen, consistent with a lower efficiency of respiratory energy coupling. In glucose-limited retentostat cultures at specific growth rates down to ~0.001 h^{-1}, both *O. parapolymorpha* strains showed high viability. Maintenance energy requirements at these extremely low growth rates were approximately 3-fold lower than estimated from faster-growing chemostat cultures, indicating a stringent-response-like behavior. Quantitative transcriptome and proteome analyses indicated condition-dependent expression patterns of Complex I subunits and of alternative NADH dehydrogenases that were consistent with physiological observations.

Importance. Since popular microbial cell factories have typically not been selected for efficient respiratory energy coupling, their ATP yields from sugar catabolism are often suboptimal. In aerobic industrial processes, suboptimal energy coupling results in reduced product yields on sugar, increased process costs for oxygen transfer, and volumetric productivity limitations due to limitations in gas transfer and cooling. This study provides insights into the contribution of mechanisms of respiratory energy coupling in the yeast cell factory *Ogataea parapolymorpha* under different growth conditions and provides a basis for rational improvement of energy coupling in yeast cell factories. Analysis of energy metabolism of *O. parapolymorpha* at extremely low specific growth rates indicated that this yeast reduces its energy requirements for cellular maintenance under extreme energy limitation. Exploration of the mechanisms for this increased energetic efficiency may contribute to an optimization of the performance of industrial processes with slow-growing eukaryotic cell factories.

Introduction

Crabtree-negative yeast species, which exhibit a respiratory sugar catabolism in aerobic batch cultures, are popular platforms for industrial production of proteins (Cereghino and Cregg 2000, Löbs *et al.* 2017b, van Dijk *et al.* 2000, Vieira Gomes *et al.* 2018, Wagner and Alper 2016). The methylotrophic, Crabtree-negative yeasts *Ogataea polymorpha* and *Ogataea parapolymorpha* (Manfrao-Netto *et al.* 2019), both formerly known as *Hansenula polymorpha*, are popular protein expression platforms because of the availability of very strong but tightly controllable, methanol-inducible promoters. They are able to consume a wide range of carbon sources and assimilate nitrate, are thermotolerant up to 50°C and exhibit fast, virtually by-product-free aerobic growth on glucose (Juergens *et al.* 2018a, Kunze *et al.* 2009, Kurtzman 2011).

Since formation of proteins and other non-dissimilatory products from sugars requires a net input of cellular energy, efficient energy coupling of respiratory sugar catabolism is an important property of microbial protein production hosts. Respiratory energy coupling can be quantified by the P/O ratio, which represents the number of moles of ATP generated per mole of oxygen atoms reduced by the respiratory chain (Hinkle 2005). In yeasts, the P/O ratio is dictated by the *in vivo* stoichiometries of electron transfer and proton translocation by respiratory chain complexes in the inner mitochondrial membrane, as well as by the stoichiometry of proton influx and ATP generation by the mitochondrial ATP synthase (Ferguson 2010). Different respiratory chain components (**Figure 4.1**) can overlap in their catalytic activities while exhibiting different stoichiometries of electron transfer and proton translocation, resulting in different P/O ratios (Kerscher 2000). Whereas the canonical machinery for transfer of electrons from ubiquinone to oxygen (cytochrome *bc1* complex or Complex III and cytochrome *c* oxidase or Complex IV) is strongly conserved among industrially relevant yeasts and fungi, major differences exist in coupling of the oxidation of mitochondrial NADH to the reduction of ubiquinone (Joseph-Horne *et al.* 2001, Kerscher *et al.* 2008, Riley *et al.* 2016).

The respiratory chains of the industrially relevant yeasts *Saccharomyces cerevisiae* and *Kluyveromyces lactis* lack the large, multisubunit proton-translocating Complex I NADH:ubiquinone oxidoreductase. Instead, they rely on a single-subunit, internal alternative NADH:ubiquinone oxidoreductase (Ndi1, generally referred to as internal alternative NADH dehydrogenase) that does not translocate protons (de Vries and Marres 1987, Hunte *et al.* 2010, Joseph-Horne *et al.* 2001, Tarrio *et al.* 2005). Other yeasts and fungi, such as *Yarrowia lipolytica* (Kerscher *et al.* 1999), exclusively utilize Complex I for respiratory oxidation of mitochondrial NADH or can express both Complex I and an internal alternative NADH dehydrogenase, e.g., *Neurospora crassa* (Duarte *et al.* 2003). *O. polymorpha* and its methylotrophic relative *Pichia pastoris*

harbor Complex I but also exhibit alternative NADH dehydrogenase activity. However, for the alternative NADH dehydrogenases, it is not known whether their catalytic sites for NADH oxidation face the mitochondrial matrix or the cytosol (Bridges *et al.* 2009, Gonzalez-Barroso *et al.* 2006) (**Figure 4.1**).

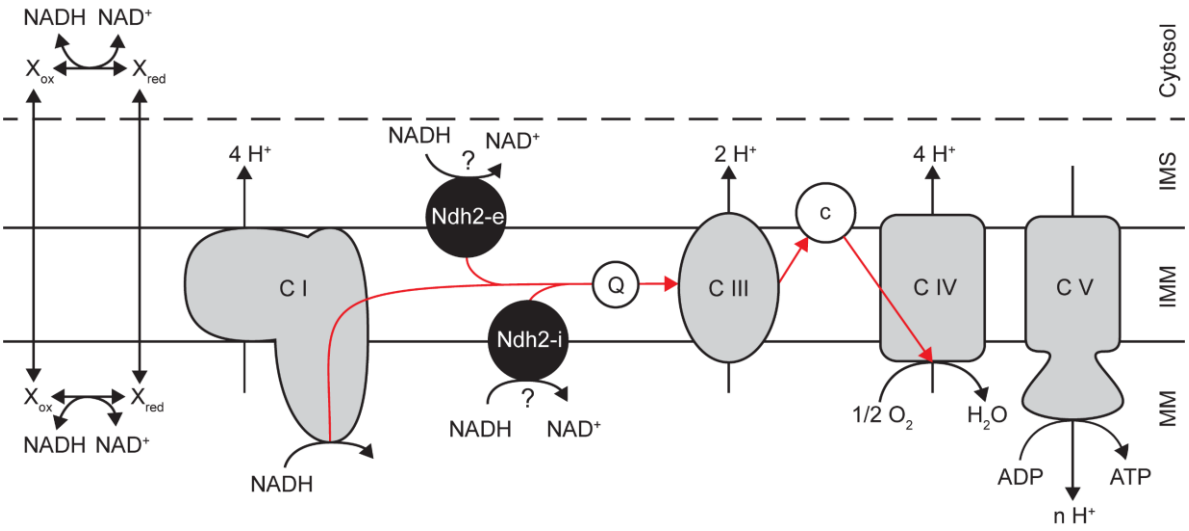


Figure 4.1: Putative respiratory chain structure of *Ogataea (para)polymorpha* illustrating routes to couple direct NADH oxidation to ATP formation. Respiratory Complex I (C I) and possibly an internal alternative NADH dehydrogenase(s) (Ndh2-i) oxidize NADH in the mitochondrial matrix (MM). NADH generated in the cytosol can be directly oxidized by an external alternative NADH dehydrogenase(s) (Ndh2-e). Shuttles, consisting of a corresponding pair of cytosolic and mitochondrial dehydrogenases, might exist which can indirectly translocate NADH over the inner mitochondrial membrane (IMM). All NADH-oxidizing respiratory enzymes donate electrons (red arrows) to the quinone pool (Q), from which they are funneled linearly through the rest of the respiratory chain, consisting of Complex III (C III), cytochrome *c* (C), and Complex IV (C IV), before reduction of oxygen to water. Contrary to many other Complex I-harboring yeasts, *O. (para)polymorpha* does not possess an alternative oxidase (Riley *et al.* 2016). Respiratory complexes I, III, and IV, but not Ndh2, contribute to the proton gradient across the inner mitochondrial membrane which is utilized by mitochondrial F_oF_1 ATPase, Complex V (C V), for formation of ATP. The dashed line represents the outer mitochondrial membrane. IMS, intermembrane space.

In organisms which can synthesize both Complex I and alternative NADH dehydrogenase(s), the relative contribution of these systems under industrially relevant conditions has not been fully elucidated. In several yeasts and fungi, studies performed with Complex I inhibitors found that there was little or no impact on specific growth rates in batch cultures or that Complex I activity was higher in late-exponential and/or stationary-phase cultures (Blondin *et al.* 1984, Bridges *et al.* 2009, Katz *et al.* 1971, Schwitzguebel and Palmer 1982), suggesting that the *in vivo* contribution of Complex I may depend on the specific growth rate and/or substrate concentration. ATP from substrate catabolism is used to meet cellular maintenance energy requirements as well as formation of new biomass (Pirt 1965). Differences in respiratory energy coupling are therefore reflected in the biomass yield on the energy substrate (which serves as the electron donor for catabolism) (Verduyn *et al.* 1991). Since respiratory oxidation of NADH via Complex I and alternative NADH dehydrogenase(s) results in different P/O ratios, the contribution of these systems can be assessed by quantitative

analysis of biomass yields on sugar and oxygen of strains in which specific systems have been inactivated. Flux balance analysis simulations indicate that, in the absence of product formation and with a constant biomass composition, exclusive use of Complex I for oxidation of mitochondrial NADH in respiratory cultures should result in a ca. 25% higher biomass yield on glucose than exclusive use of internal alternative NADH dehydrogenase (Correia *et al.* 2017).

Biomass yields and maintenance energy requirements of the yeasts *S. cerevisiae* and *P. pastoris* were previously studied under extreme glucose limitation in retentostat cultures (Boender *et al.* 2009, Rebnegger *et al.* 2016, Vos *et al.* 2016). Retentostats are continuous cultivation setups with biomass retention in which biomass-specific rates of energy substrate consumption are progressively decreased until they eventually fulfill maintenance energy requirements of only metabolically active cells (Ercan *et al.* 2015). Maintenance requirements include energy expenses for macromolecular turnover, membrane gradient homeostasis, and protective measures such as detoxification of reactive oxygen species (ROS) (van Bodegom 2007).

The aim of this study is to quantitatively assess the relative contribution of Complex I and the alternative NADH dehydrogenases to respiratory oxidation of NADH, to biomass yields, and to maintenance energy metabolism in *O. parapolymorpha*. To this end, a Complex I-deficient strain was constructed by disruption of the structural gene for the essential Nubm (51 kDa) subunit. The physiology of the mutant strain was then analyzed in aerobic, glucose-grown batch, chemostat, and retentostat cultures in bioreactors, covering a range of specific growth rates, and results were compared to data obtained with the congenic wild-type reference strain. In addition to quantitative analyses of biomass yields and maintenance energy requirements, strain- and condition-dependent adaptations were investigated by transcriptome and proteome analysis.

Results

Disruption of Complex I has a negligible effect on growth physiology in aerobic, glucose-grown batch cultures. To assess under which cultivation conditions *O. parapolymorpha* utilizes Complex I, a strain devoid of Complex I activity was constructed. The structural gene encoding the essential, nuclearly encoded Nubm (51 kDa) subunit was disrupted in wild-type *O. parapolymorpha* strain CBS11895 by CRISPR/Cas9-assisted introduction of a single-nucleotide frameshift. Since Nubm, which is part of the peripheral N-module of Complex I, contains the NADH-binding pocket (Fecke *et al.* 1994), the frameshift mutation was expected to abolish Complex I-mediated NADH oxidation and block entry of NADH-derived electrons into the enzyme. Whole-genome sequencing of the resulting strain, *O. parapolymorpha* IMD010, did not reveal any additional mutations in coding sequences.

For an initial physiological comparison, strain IMD010 and its congenic wild-type strain *O. parapolymorpha* CBS11895 were grown on glucose in aerobic bioreactor batch cultures. Under these glucose-excess conditions, both strains exhibited a specific growth rate of 0.37 h^{-1} (**Table 4.1**) and respiratory metabolism with negligible production of extracellular pyruvate (<0.1 mM), citrate (<0.1 mM), and ethanol (<1 mM). Biomass-specific glucose uptake rates and biomass yields of the two strains were not significantly different (**Table 4.1**). These observations indicated that Complex I does not significantly contribute to NADH oxidation by *O. parapolymorpha* during aerobic batch cultivation on glucose.

Table 4.1: **Physiology of wild-type strain *O. parapolymorpha* CBS11895 and Complex I-disrupted mutant IMD010 in aerobic glucose-grown bioreactor batch cultures.**

Parameter[a]	Value for the parameter in:[b]		*P* value[c]
	CBS11895	IMD010	
μ (1/h)	0.37 ± 0.01	0.37 ± 0.00	0.78
$q_{Glucose}$ (mmol glucose/g biomass)/h	−3.88 ± 0.02	−4.01 ± 0.00	0.11
$Y_{X/S}$ (g biomass/g glucose)	0.53 ± 0.01	0.52 ± 0.00	0.21

a: μ, specific growth rate based on measurements of biomass dry weight concentration; $q_{Glucose}$, biomass-specific glucose uptake rate during exponential growth phase; $Y_{X/S}$, yield of biomass dry weight on glucose during exponential growth phase.

b: Reported values are means ± standard errors of the means calculated from two independent cultures where errors smaller than the number of reported digits are rounded to 0.

c: Reported *P* values (Student's *t* test) refer to the difference between mean values observed for CBS11895 and those for IMD010.

Inactivation of Complex I decreases biomass yields on glucose and oxygen in glucose-limited chemostat cultures. To investigate the contribution of Complex I at lower specific growth rates and at growth-limiting concentrations of glucose, the wild-type *O. parapolymorpha* strain CBS11895 and the Complex I-deficient strain IMD010 were grown in aerobic, glucose-limited chemostat cultures (**Table 4.2**). At a dilution rate of 0.1 h^{-1}, steady-state cultures of both strains exhibited a fully respiratory metabolism, without significant by-product formation and with residual glucose concentrations below 10 µM. Biomass protein contents of strains CBS11895 and IMD010 were 0.40 and 0.42 g of protein/g of biomass, respectively. Since the protein fraction of biomass accounts for the majority of the energetic costs of biosynthesis (Verduyn 1991), this result indicates that the two strains exhibit similar energetic requirements for biomass formation ($Y_{X/ATP}$).

Table 4.2: Physiology of wild-type *O. parapolymorpha* strain CBS11895 and Complex I-disrupted mutant IMD010 in aerobic glucose-limited chemostat cultures at a dilution rate of 0.1 h^{-1}.

Parameter[a]	Value for the parameter in:[b]		*P* value[c]
	CBS11895	IMD010	
Actual dilution rate (1/h)	0.099 ± 0.000	0.100 ± 0.001	
Reservoir glucose (g/L)	7.37 ± 0.02	7.40 ± 0.00	
Residual glucose (mM)	BDL	BDL	
$Y_{X/S}$ (g biomass/g glucose)	0.51 ± 0.00	0.43 ± 0.00	0.003
$Y_{X/O2}$ (g biomass/g O_2)	1.26 ± 0.01	0.88 ± 0.01	0.002
RQ	1.03 ± 0.00	1.03 ± 0.01	0.91
$q_{Glucose}$ (mmol/g biomass)/h	−1.08 ± 0.01	−1.27 ± 0.01	0.005
q_{CO2} (mmol/g biomass)/h	2.52 ± 0.02	3.64 ± 0.04	0.005
q_{O2} (mmol/g biomass)/h	−2.44 ± 0.03	−3.54 ± 0.06	0.013
C_X (g biomass/liter)	3.73 ± 0.02	3.21 ± 0.01	0.009
Protein content (g protein/g biomass)	0.40 ± 0.01	0.42 ± 0.00	0.22
Cell viability (%)	99.6 ± 0.0	99.9 ± 0.0	0.011
Carbon recovery (%)	99.5 ± 0.2	99.9 ± 0.4	0.53

a: $Y_{X/S}$ and $Y_{X/O2}$, yield of biomass dry weight on glucose and oxygen, respectively; RQ, respiratory quotient; $q_{Glucose}$, q_{CO2}, and q_{O2}, biomass-specific uptake/production rates of glucose, CO_2, and O_2, respectively; C_X, biomass dry weight concentration.

b: Reported values are means ± standard errors of the means calculated from two independent steady-state cultures where errors smaller than the number of reported digits are rounded to 0. BDL, below the detection limit (10 µM). Reported cell viability is based on propidium-iodide staining. Carbon recovery calculations are based on a biomass carbon content of 48% (wt/wt).

c: Reported *P* values (Student's *t* test) refer to the difference between mean values observed for CBS11895 and those for IMD010.

In agreement with earlier observations (Juergens *et al.* 2018a), the wild-type strain showed a biomass yield of 0.51 g of biomass/g of glucose. A 16% lower biomass yield on glucose for strain IMD010 indicated that, in the absence of a functional Complex I, a larger fraction of the substrate needed to be respired to generate the same amount of ATP. Strain IMD010 also exhibited a 30% lower biomass yield on oxygen (**Table 4.2**), which, under the assumption of similar $Y_{X/ATP}$ values and negligible maintenance energy requirements in these cultures, is equivalent to a 30% lower *in vivo* P/O ratio (Katz *et al.* 1971). Considering that alternative NADH dehydrogenases do not translocate protons and therefore are expected to conserve 40% less ATP than Complex I (Hinkle 2005, Kerscher *et al.* 2008), the 30% lower *in vivo* P/O ratio is consistent with the phenotype expected when oxidation of mitochondrial NADH from glucose catabolism occurs via Complex I in strain CBS11895 and is replaced by oxidation via alternative NADH dehydrogenase(s) in strain IMD010. Similar differences between strains CBS11895 and IMD010 were observed at a dilution rate of 0.025 h^{-1} (see **Table S4.1** in the supplemental material).

***O. parapolymorpha* decreases its maintenance energy requirements at near-zero growth rates in retentostat cultures, independent of Complex I.** Based on retentostat experiments with *S. cerevisiae*, maintenance energy requirements of yeasts were initially assumed to be growth rate independent (Boender *et al.* 2009, Vos *et al.* 2016). This conclusion was in marked contrast with observations on several bacteria, in which a stringent response leads to decreased maintenance energy requirements at very low specific growth rates (Arbige and Chesbro 1982, Chesbro *et al.* 1979, Tappe *et al.* 1999). Recent experiments on aerobic, glucose-limited retentostats of the Crabtree-negative yeast *P. pastoris* showed that, similarly, maintenance energy requirements at near-zero growth rates were approximately 3-fold lower than predicted from data obtained at higher specific growth rates (Rebnegger *et al.* 2016). While these results indicate a stringent-response-like adaptation of non-*Saccharomyces* yeasts at near-zero growth rates, it is unclear whether this is related to their expression of a functional Complex I NADH dehydrogenase. *O. parapolymorpha*, like *P. pastoris*, harbors both Complex I and alternative NADH dehydrogenases, and we tested if a similar modulation of maintenance energy requirements occurred in *O. parapolymorpha* and if this was partly caused by a redistribution of respiratory flux to Complex I.

Maintenance energy requirement (m_S) and maximum theoretical biomass yield (Y_{xs}^{max}) (Pirt 1965) values of *O. parapolymorpha* strains CBS11895 and IMD010 were first estimated from biomass-specific glucose uptake rates of aerobic, glucose-limited chemostat cultures grown at 0.025 h^{-1} and 0.1 h^{-1}. Consistent with a lower energetic efficiency of the Complex I-deficient strain IMD010, its estimated m_S was higher (0.0241 ± 0.0008 versus 0.0142 ± 0.0008 [g glucose/g biomass]/h) and its Y_{xs}^{max} was

lower (0.485 ± 0.003 versus 0.545 ± 0.003 g biomass/g of glucose) than values obtained with the wild-type strain CBS11895 (**Figure S4.1**).

During 23 days of retentostat cultivation, specific growth rates of strains CBS11895 and IMD010 decreased to approximately 0.001 h^{-1}, corresponding to a doubling time of over 28 days. As biomass concentrations in the retentostats increased and the specific growth rate progressively decreased, culture viability, measured by propidium iodide (PI) staining and CFU counts, remained near 100% (**Figure 4.2; S4.2**). At near-zero growth rates, the biomass protein content of CBS11895 decreased to 0.33 g protein/g of biomass, while biomass composition of strain IMD010 remained the same as that observed in chemostat cultures at higher specific growth rates (**Figure S4.3**). In the retentostats, the biomass-specific glucose uptake rates (q_S) of both strains decreased below the m_S values estimated from chemostat data, reaching 0.0062 ± 0.0001 and 0.0081 ± 0.0002 (g glucose/g biomass)/h at the end of retentostat cultivation for strains CBS11895 and IMD010, respectively (**Figure 4.3**).

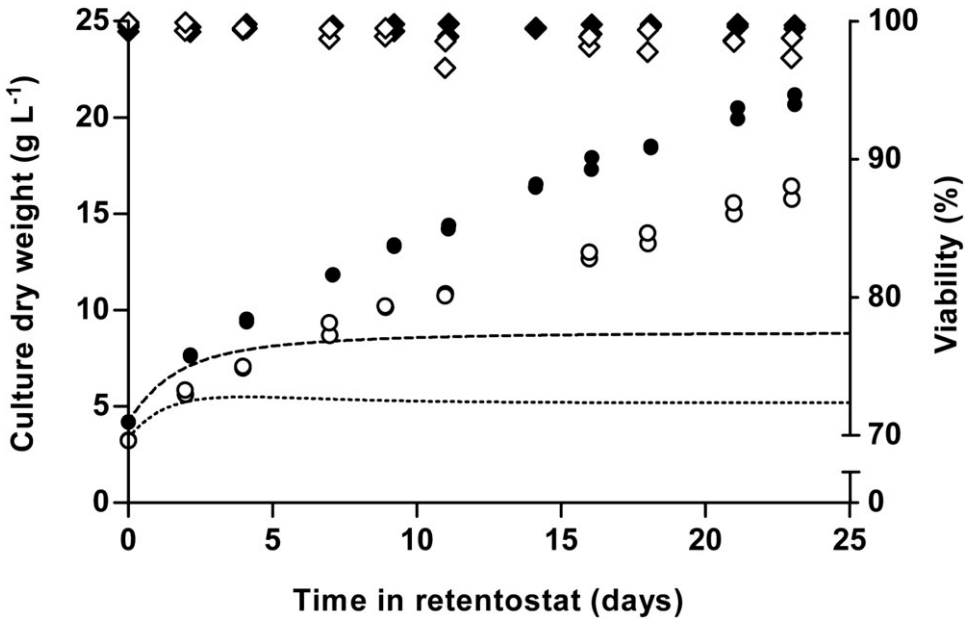


Figure 4.2: Biomass accumulation profile and viability of aerobic, glucose-limited retentostat cultures of wild-type *O. parapolymorpha* CBS11895 and the congenic Complex I-deficient strain IMD010. The retentostat phase was initiated from steady-state chemostat cultures (D = 0.025 h^{-1}) at time zero. Depicted are the measured biomass dry weight concentrations (circles) and culture viability based on propidium iodide staining (diamonds) of two independent cultures each of strains CBS11895 (closed symbols) and IMD010 (open symbols), as well as the predicted biomass accumulation profiles of CBS11895 (dashed line) and IMD010 (dotted line) based on m_S and $Y_{X/S}^{max}$ values estimated from chemostat cultures grown at 0.1 and 0.025 h^{-1}. The mean biomass concentration of CBS11895 was significantly higher than of that of IMD010 at each equivalent sampling point (Student's t test, $P < 0.05$).

By performing linear regression on sections of adjacent sample points, m_S values were estimated throughout the retentostat cultivation runs from μ and q_S values (**Figure 4.4**). From this analysis, m_S values of both strains were found to decrease during retentostat cultivation to values that were 2.5- to 3-fold lower than those estimated from chemostat data (**Figure 4.4**). In accordance with an m_S that was lower than expected, biomass concentrations of CBS11895 during the retentostat cultivation increased above the maximum level predicted based on a growth rate-independent m_S (**Figure 4.2**). These observations demonstrated that at low specific growth rates, similar to *P. pastoris*, *O. parapolymorpha* exhibits a growth rate-dependent substrate requirement for maintenance. Since the Complex I-deficient *O. parapolymorpha* strain IMD010 exhibited a decrease in maintenance energy requirements similar to that of the wild-type strain CBS11895 (**Figure 4.2**), this adaptation is independent of the contribution of Complex I to respiratory energy coupling.

Transcriptional adaptations of *O. parapolymorpha* to lack of functional Complex I. To investigate whether the absence of the Complex I Nubm subunit results in transcriptional adaptations in *O. parapolymorpha*, transcriptome sequencing (RNA-Seq) was performed on samples taken from the glucose-grown bioreactor batch cultures, the chemostat cultures grown at dilution rates of 0.1 and 0.025 h^{-1}, and from the late-stage retentostat cultures (samples taken after 23 days, at a specific growth rate of ca. 0.001 h^{-1}) of strains CBS11895 and IMD010.

To focus on large changes in expression levels, genes were considered significantly differentially expressed between strains CBS11895 and IMD010 when the absolute $\log_2$ fold change of expression was larger than 2 and the false-discovery rate (FDR) was below 0.001. Whereas large differences in growth phenotypes were observed in glucose-limited chemostat and retentostat cultures, the absence of functional Complex I only marginally affected the transcriptome of IMD010 grown under these conditions (**Figure 4.5B**). However, while CBS11895 and IMD010 exhibited the same growth phenotypes in glucose-grown batch cultures, a large number of genes were significantly differentially expressed between the strains under these conditions, and most of these showed higher transcript levels in the Complex I-deficient strain IMD010 (**Figure 4.5A**). These 419 differentially expressed genes did not contain any of the known subunits of respiratory Complex I. Among the set of 275 of the 410 upregulated genes for which an *S. cerevisiae* orthologue could be identified, the Gene Ontology (GO) terms related to organization and biogenesis of cellular components were enriched (**Figure 4.5A**; see **Table S4.2** for an extended list of enriched GO terms).

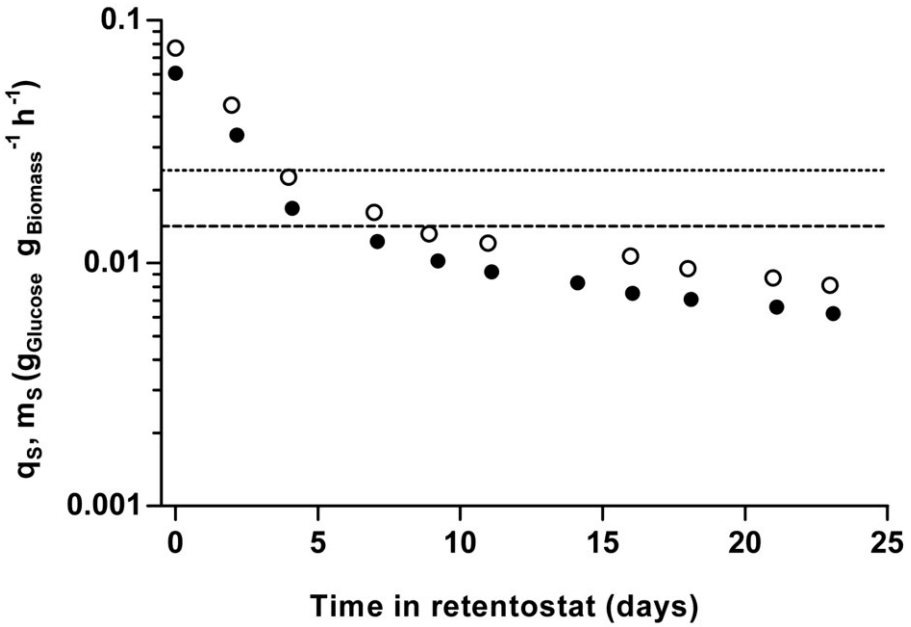


Figure 4.3: Biomass-specific glucose uptake rates (q_S) during aerobic glucose-limited retentostat cultivation of wild-type *O. parapolymorpha* CBS11895 and the congenic Complex I-deficient strain IMD010. Depicted q_S values are means ± standard errors of the means (error bars smaller than symbol size) of two independent retentostat cultures each of CBS11895 (closed circles) and IMD010 (open circles) and were directly calculated from biomass accumulation. The values plotted at time zero correspond to the q_S value in the steady-state chemostat cultures at 0.025 h^{-1} that preceded the retentostat cultures. Horizontal lines indicate the maintenance energy requirements (m_S) calculated from chemostat cultures grown at 0.1 and 0.025 h^{-1} for strains CBS11895 (dashed) and IMD010 (dotted). With the exception of the values calculated at 7 and 9 days of cultivation, the q_S values of CBS11895 were significantly lower than those of IMD010 at each equivalent sampling point (Student's *t* test, $P < 0.05$).

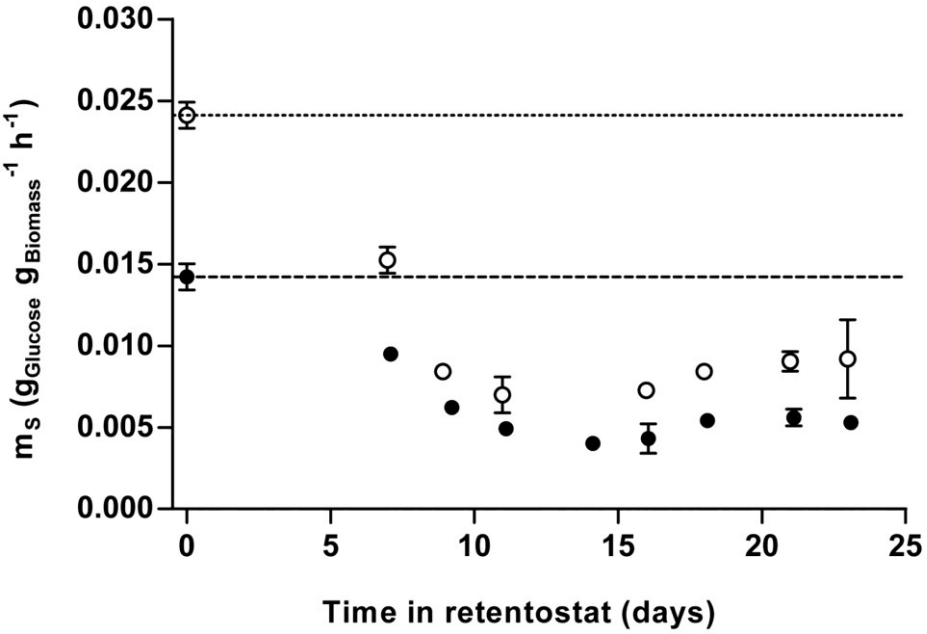


Figure 4.4: Depicted m_S values are means ± standard errors of the means of two independent retentostat cultures each of strains CBS11895 (closed circles) and IMD010 (open circles) and were calculated via linear regression from sets of corresponding μ and q_S values (directly calculated from biomass accumulation) from five adjacent sample points. The values plotted at time zero correspond to m_S values determined by chemostat cultivation at 0.1 and 0.025 h^{-1}, also represented by dashed (CBS11895) and dotted (IMD010) lines. With the exception of time points at 16 (CBS11895) as well as at 7 and 23 (IMD010) days of cultivation, m_S values of both strains were found significantly lower than those determined by chemostat cultivation (Student's t test, $P < 0.05$).

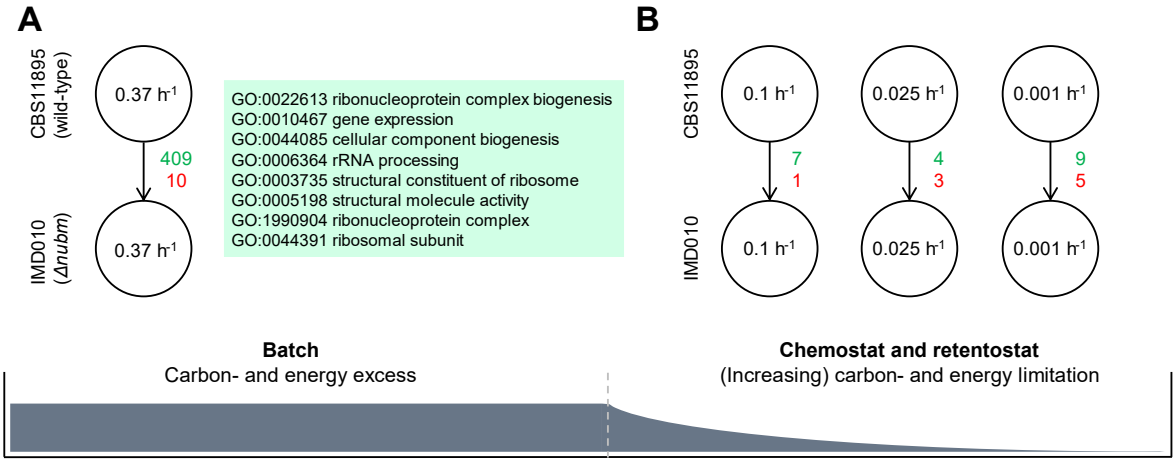


Figure 4.5: Transcriptional response of *O. parapolymorpha* to a lack of functional respiratory Complex I. Green (upregulated) and red (downregulated) numbers indicate how many genes were found significantly differentially expressed (absolute $\log_2$ fold change of >2; FDR < 0.001) in strain IMD010 (disrupted Complex I Nubm subunit) compared to levels in strain CBS11895 (wild type) in glucose-grown batch cultures **(A)** and glucose-limited chemostat (0.1 and 0.025 h^{-1}) and late-stage retentostat cultures (0.001 h^{-1}) **(B)**. Boxed in green are the most highly enriched GO terms within the set of upregulated genes in IMD010 under batch conditions (based on 275 out of 409 genes for which an *S. cerevisiae* ortholog could be identified) (see Table S4.2 in the supplemental material for an extended list). Numbers inside circles represent the specific growth rate/dilution rate of the different cultures

Absence of functional Complex I also affected transcriptional adaptations to increasingly lower specific growth rates in glucose-limited cultivation regimes. A total of 1,699 and 1,074 genes whose transcript levels correlated positively or negatively with the specific growth rate in glucose-limited cultures were identified for strains CBS11895 and IMD010, respectively (**Figure 4.6A**). For the majority of these genes, transcript levels correlated positively with the specific growth rate; i.e., expression was lower with low specific growth rates, in both CBS11895 and IMD010 (**Figure 4.6A**). Approximately two-thirds of the identified growth rate-correlated genes for strain IMD010 exhibited the same regulation in strain CBS11895, whereas the majority of the genes identified to be growth rate correlated in CBS11895 were specifically identified in this strain background (**Figure 4.6A**). The response to the low specific growth rates of both CBS11895 and IMD010 was characterized by the transcriptional downregulation of genes involved in biosynthesis and metabolic processes as indicated by the enrichment of GO terms "biosynthetic process", "cellular amino acid metabolic process", and "catalytic activity" in this gene set (231, 54, and 263 out of 414 genes, respectively) (**Figure 4.6B**; see **Table S4.3** for an extended list of enriched GO terms). Among the genes that were positively correlated with the specific growth rate only in strain CBS11895, an enrichment of GO terms related to chromosome organization, DNA binding, and the cytoskeleton was observed, while the GO term "catalytic activity, acting on a tRNA" was enriched in the set of positively correlated genes unique to strain IMD010 (**Figure 4.6B**; **Table S4.3**).

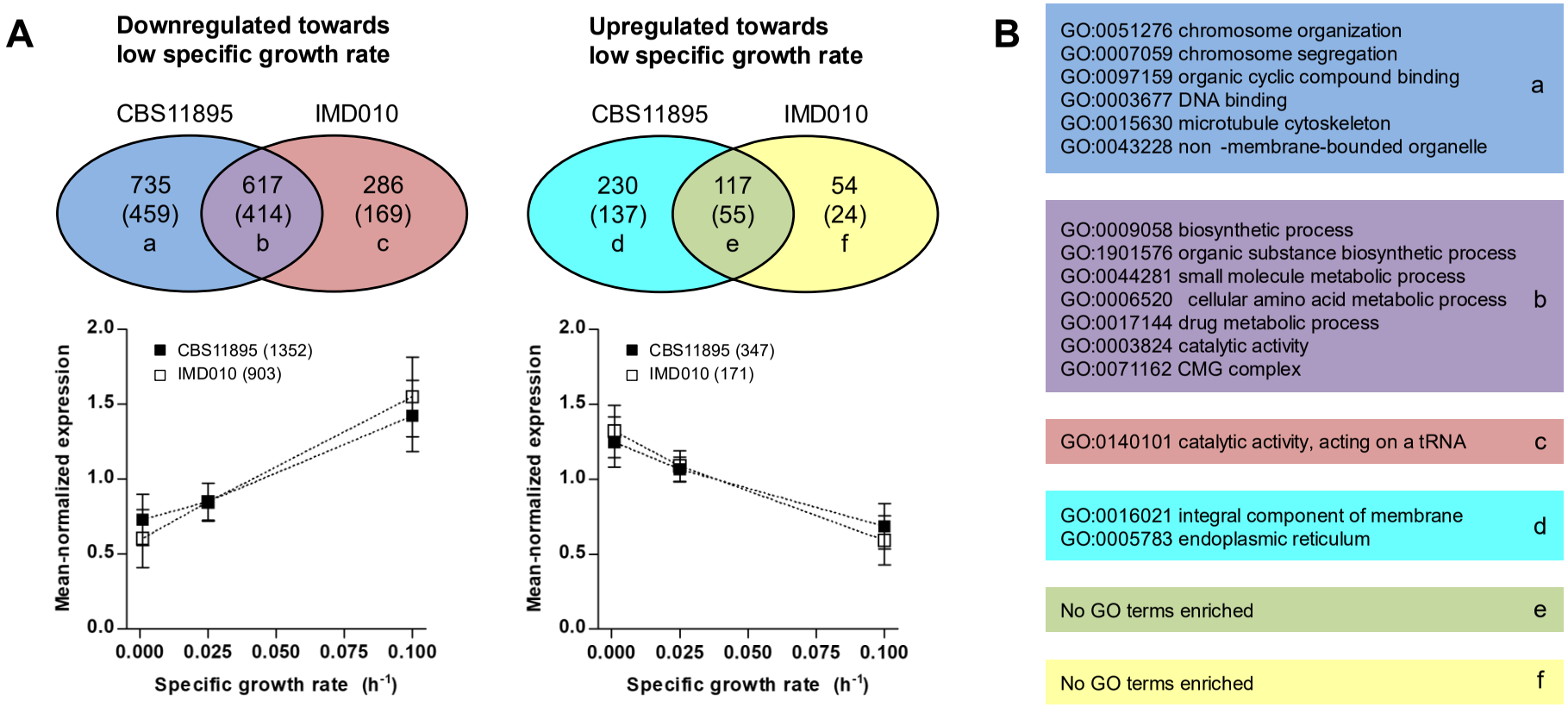


Figure 4.6: Transcriptional adaptation of wild-type *O. parapolymorpha* CBS11895 and congenic Complex I-deficient strain IMD010 to increasingly lower specific growth rates. (A) The bottom graphs show mean-normalized expression levels of genes identified to be positively (left) and negatively (right) correlated with specific growth rates in strains CBS11895 and IMD010, based on samples taken from glucose-limited chemostat (0.1 and 0.025 h^{-1}) and late-stage retentostat (0.001 h^{-1}) cultures (data presented as means ± standard deviation). Venn diagrams at the top indicate the overlap between genes identified to be positively (left) and negatively (right) correlated with specific growth rates identified for CBS11895 and IMD010. Numbers in parentheses indicate genes for which an *S. cerevisiae* ortholog could be identified. **(B)** Significantly enriched GO terms identified in the sets of genes with growth rate-correlated expression. Colors and lowercase letters correspond to Venn diagrams in panel A. For each set, the two most highly enriched GO terms of a category (biological process, molecular function, and cellular component) are listed, except for set b, for which all significantly enriched GO terms are shown (see Table S4.3 in the supplemental material for extended list).

Finally, among the smaller sets of genes which exhibited negative correlation of transcript levels with the specific growth rate in glucose-limited cultures, an enrichment of GO terms was detected only for the set of genes uniquely regulated in strain CBS11895 (related to integral membrane components and the endoplasmic reticulum) (**Figure 4.6B; Table S4.3**).

Condition-dependent redistribution of respiratory fluxes between Complex I and alternative mechanisms for NADH (re)oxidation. In glucose-limited chemostat cultures, the Complex I-deficient strain *O. parapolymorpha* IMD010 exhibited a lower biomass yield on substrate and oxygen than the wild-type strain CBS11895 but retained a fully respiratory metabolism. These observations indicated that glucose-limited cultures of strain IMD010 employed alternative, energetically less efficient mechanism(s), such as alternative NADH dehydrogenase(s), to replace the role of Complex I in NADH oxidation. Based on sequence homology to known alternative NADH dehydrogenases and the C-terminal domain unique to this class of enzymes (Feng *et al.* 2012), the genome of *O. parapolymorpha* was predicted to encode three alternative NADH dehydrogenases (**Figure S4.4**), here referred to as Ndh2-1, Ndh2-2, and Ndh2-3 (encoded by HPODL_02792, HPODL_00256, and HPODL_02018, respectively). Depending on substrate specificity and localization on the inner mitochondrial membrane, each of these enzymes could potentially contribute to reoxidation of NADH generated in the mitochondrial matrix.

To investigate condition-dependent expression of these alternative dehydrogenases, their protein abundance levels and those of Complex I subunits were determined by mass spectrometry (MS)-based proteomics analysis on samples taken from the glucose-grown batch, chemostat (dilution rate of 0.1 and 0.025 h^{-1}), and late-stage retentostat cultures (specific growth rate of ca. 0.001 h^{-1}) for strains CBS11895 and IMD010. Proteomics analysis of these samples detected 1,351 *O. parapolymorpha* proteins with high combined detection confidence (FDR of <1%), including the three alternative NADH dehydrogenases as well as nearly all subunits of Complex I (see Data Set S1 in the supplementary material for protein abundance data).

Mean-normalized transcript and protein abundance levels of the three alternative NADH dehydrogenases were compared to those of the seven essential nuclearly encoded subunits of Complex I to investigate their strain- and condition-dependent expression (**Figure 4.7**). In strains CBS11895 and IMD010, transcript levels of the seven Complex I subunits were on average 2.3- and 2.6-fold lower, respectively, in glucose-grown batch cultures than in chemostat cultures grown at a dilution rate of 0.1 h^{-1}.

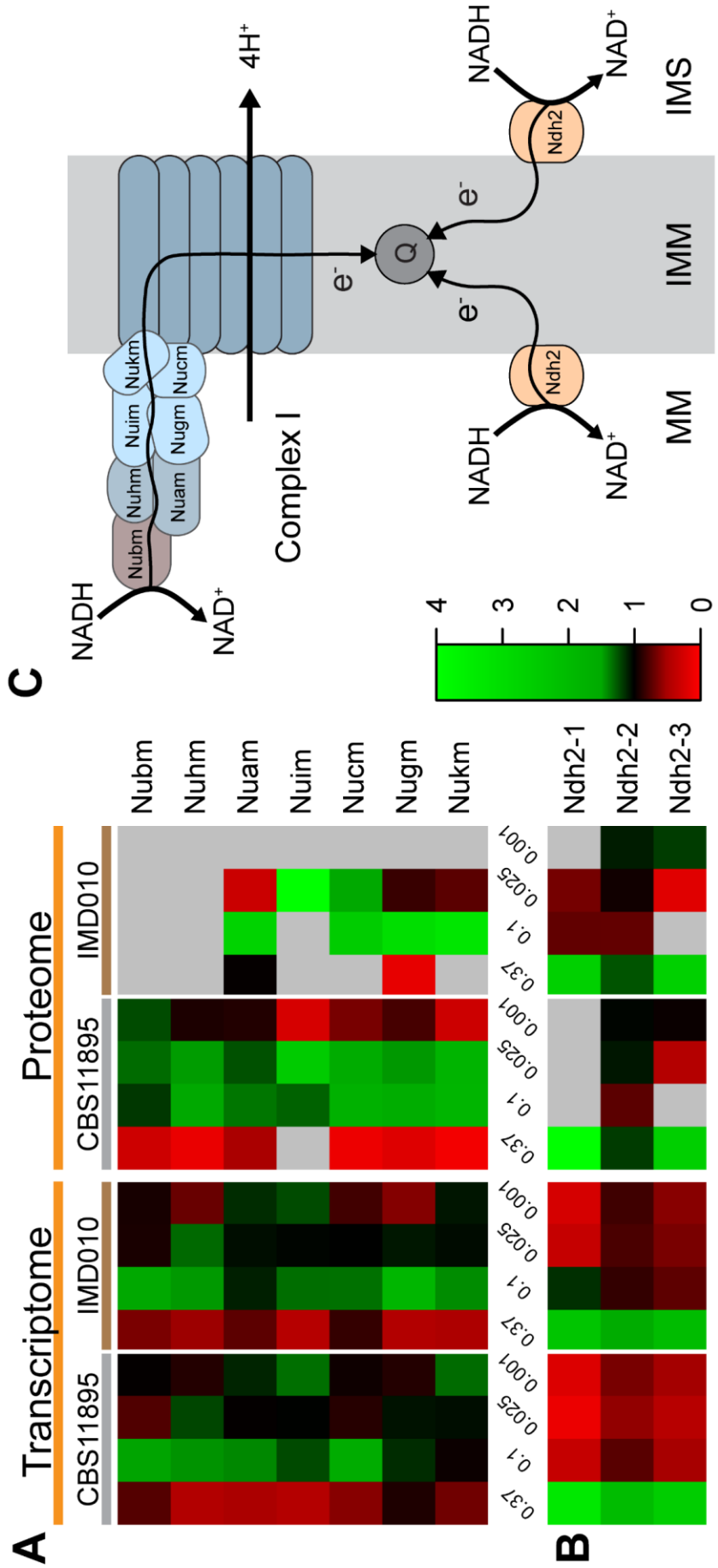


Figure 4.7: Mean-normalized transcript and protein abundances of essential Complex I subunits (A) and alternative NADH dehydrogenases (B) in *O. parapolymorpha* strains CBS11895 (wild type) and IMD010 (disrupted Complex I Nubm subunit). Samples were taken from duplicate independent aerobic, glucose-grown batch (0.37 h^{-1}), chemostat (0.1 and 0.025 h^{-1}), and late-stage retentostat (0.001 h^{-1}) cultures. Transcript and protein abundances were mean normalized separately for each gene and strain. Gray, protein not detected based on criteria described in Materials and Methods. **(C)** The location and catalyzed reactions of the enzymes. Localization of the three alternative NADH dehydrogenases is unknown, and any of the enzymes could be internally (MM) or externally (IMS) localized. MM, mitochondrial matrix; IMM, inner mitochondrial membrane; IMS, intermembrane space.

Differences in protein levels of these subunits were more pronounced. In batch cultures of strain CBS11895, these subunits were less abundantly detected, and protein levels were on average 11.3-fold lower than in the chemostat cultures while most subunits were not detected at all in batch cultures of strain IMD010 (**Figure 4.7; Table S4.4**). In addition to the disrupted Nubm subunit, the Nuhm (24 kDa) subunit of Complex I was not detected in any of the proteome analyses on strain IMD010. In contrast to the changes in expression of Complex I, in both *O. parapolymorpha* strains the three alternative NADH dehydrogenases Ndh2-1, Ndh2-2 and Ndh2-3 consistently showed higher transcript and protein levels in batch cultures than in chemostat cultures grown at 0.1 h^{-1}. (**Figure 4.7**). These observations are consistent with the similar growth characteristics of strains CBS11895 and IMD010 in glucose-grown batch cultures and support the conclusion that Complex I plays an insignificant role in respiratory NADH reoxidation by *O. parapolymorpha* at high glucose concentrations.

Transcript levels of Complex I subunits in retentostat cultures were similar to those in chemostat cultures, while the corresponding proteins were less abundantly detected and exhibited lower protein levels. In strain CBS11895, protein levels of the represented essential Complex I subunits were on average 3.5-fold lower in late-stage retentostats than in chemostat cultures grown at a dilution rate of 0.025 h^{-1}. In strain IMD010, most Complex I subunits, including the numerous accessory subunits, were not detected under these conditions (**Figure 4.7**, **Table S4.4**).

Discussion

Contribution of Complex I to respiratory glucose metabolism and growth energetics. Physiological analysis of *O. parapolymorpha* CBS11895 and its congenic mutant IMD010 showed that the Complex I NADH dehydrogenase does not play a major role in respiratory NADH oxidation in aerobic, glucose-grown batch cultures. The reported insensitivity to Complex I inhibitors of respiratory rates of aerobic, glucose-grown batch cultures of *Candida utilis* and *Dekkera bruxellensis* (Blondin *et al.* 1984, Katz *et al.* 1971) suggests that the physiological role and regulation of Complex I in these facultatively fermentative yeasts resemble that in *O. parapolymorpha*. In contrast, Complex I is essential for growth of the respiratory yeast *Yarrowia lipolytica* (Kerscher 2000) while in the filamentous fungi *Neurospora crassa* and *Aspergillus niger*, its absence negatively affects specific growth rates and/or biomass yields in aerobic batch cultures (Fecke *et al.* 1994, Prömper *et al.* 1993).

While aerobic, glucose-limited chemostat cultures of the Complex-I deficient strain IMD010 showed lower biomass yields on glucose and oxygen than the reference strain CBS11895, they still exhibited a fully respiratory metabolism. Clearly, another mechanism for oxidation of mitochondrial NADH, with a lower ATP yield from oxidative phosphorylation, compensated for the absence of a functional Complex I in these cultures. Since NADH cannot permeate the inner mitochondrial membrane (von Jagow and Klingenberg 1970), *O. parapolymorpha* additionally requires a mechanism for respiratory oxidation of NADH generated in the cytosol by glycolysis. By analogy of the situation in *S. cerevisiae* and other fungi (Antos-Krzeminska and Jarmuszkiewicz 2019, Bakker *et al.* 2001), it therefore seemed likely that the *O. parapolymorpha* genome encodes external alternative NADH dehydrogenase in combination with an NADH shuttle mechanism and/or a matrix-oriented (internal) alternative NADH dehydrogenase. For example, an ethanol-acetaldehyde shuttle in *S. cerevisiae* has been shown to shuttle electrons from mitochondrial NADH to the cytosol (Bakker *et al.* 2000). Such a shuttle requires cytosolic and mitochondrial isoenzymes of NAD-linked alcohol dehydrogenase, which are both present in *O. parapolymorpha* (Suwannarangsee *et al.* 2012, Suwannarangsee *et al.* 2010).

Transcript levels of a gene encoding a putative alternative NADH dehydrogenase (Ndh2-1) were significantly higher in glucose-limited chemostat cultures (dilution rate of 0.1 h^{-1}) of *O. parapolymorpha* IMD010 than in the wild type (log_2 fold change of 2.2). Moreover, Ndh2-1 was detected in the proteome of strain IMD010 grown under these conditions but not in that of strain CBS11895. These observations are consistent with Ndh2-1 being an internal, non-proton-translocating NADH dehydrogenase that can compensate for the absence of functional Complex I in the mutant strain IMD010 and in batch cultures of the wild-type strain CBS11895. Coexistence of internal alternative

NADH dehydrogenase and Complex I has been observed in other fungi, including *N. crassa* (Duarte *et al.* 2003), *C. utilis* (Katz *et al.* 1971), and *A. niger* (Wallrath *et al.* 1991).

Proteome and transcriptome analysis of respiratory Complex I subunits and alternative NADH dehydrogenases in *O. parapolymorpha* CBS11895 indicated a switch from energy-efficient respiration via Complex I in glucose-limited chemostat cultures to less efficient respiration via alternative NADH dehydrogenases in fast-growing batch cultures. It has been previously suggested that respiratory NADH oxidation by a simple single-subunit NADH dehydrogenase instead of the large multisubunit Complex I may be beneficial when energy substrate is abundantly available (Kerscher 2000). Expressed per amount of protein, alternative NADH dehydrogenases are likely to allow for faster oxidation of NADH than the multisubunit Complex I. A switch to the energetically less efficient alternative dehydrogenases is therefore consistent with a strategy in which metabolic rates are maximized under substrate excess while energy substrate limitation is coupled to optimization of energy efficiency (Molenaar *et al.* 2009, Pfeiffer *et al.* 2001). Similar trade-offs involving pathways with a high ATP yield but high protein cost (i.e., low protein efficiency) have been implicated in overflow metabolism in *Escherichia coli*, *S. cerevisiae*, and human muscle cells (Chen and Nielsen 2019, Nilsson *et al.* 2019).

Analysis of late-stage retentostat cultures suggested a higher utilization of alternative NADH dehydrogenases at extremely low specific growth rates than in faster chemostat cultures. Condition-dependent use of the different NADH dehydrogenases has been suggested to reflect the need to balance energy demand with NAD^+ regeneration (Kerscher *et al.* 2008, Marreiros *et al.* 2016b) and to prevent reactive oxygen species (ROS) formation, potentially by altering the degree of reduction of the quinone pool (Antos-Krzeminska and Jarmuszkiewicz 2019, Dominiak *et al.* 2018). Alternative, less efficient respiratory pathways are widespread in yeast (Guerrero-Castillo *et al.* 2011, Marreiros *et al.* 2016b, Veiga *et al.* 2003), and examples exist of species that redirect flux through these pathways under low-energy-substrate conditions to limit ROS formation (Cabrera-Orefice *et al.* 2014, Guerrero-Castillo *et al.* 2012). Apparently, in these organisms decreased ROS formation can outweigh the benefits of increased energetic efficiency under these conditions, and a similar mechanism could be beneficial for long-term survival in *O. parapolymorpha* at very low specific growth rates as well. Indeed, investigation of the expression of catalase as well as candidate genes for superoxide dismutase and cytochrome c peroxidase demonstrated that out of the five genes for which protein abundances could reliably be detected, four were more abundantly expressed at low specific growth rates (see **Figure S4.5** in the supplementary material).

Regulation of Complex I expression. Under several of the tested conditions, many essential Complex I subunits were not detected in proteome analyses on the mutant strain IMD010. In this strain, the Nuhm (24 kDa) subunit was not even detected in chemostat samples in which other Complex I subunits were most abundant. In normal Complex I assembly, Nubm and Nuhm, which are both part of the Complex I N-module, preassemble into a stable heterodimer which is retained even after breakdown of Complex I after inhibition of mitochondrial translation (Guerrero-Castillo *et al.* 2017). Destabilization of Nuhm caused by deletion of the *NUBM* gene would resemble observations on human Complex I, which showed that loss of individual subunits affected protein abundance of other subunits from the same structural module (Stroud *et al.* 2016).

In mammals, plants, and fungi such as *Y. lipolytica* and *N. crassa*, Complex I forms respiratory supercomplexes with complexes III and IV (Eubel *et al.* 2004, Marques *et al.* 2007, Nubel *et al.* 2009, Schägger and Pfeiffer 2000). A study on mammalian cells did not detect formation of supercomplexes with a Complex I lacking the Nubm-containing N-module (Guerrero-Castillo *et al.* 2017). However, the fast respiratory growth of strain IMD010 indicated that, even in the complete absence of the Nubm subunit, *O. parapolymorpha* expressed a functional respiratory chain. If required for respiration (Hirst 2018), respiratory supercomplexes in strain IMD010 might resemble the supercomplex-like structures observed in *S. cerevisiae*, which are comprised of complexes III and IV with the internal alternative NADH dehydrogenase Ndi1 (Matus-Ortega *et al.* 2015).

Maintenance energy requirements in *O. parapolymorpha*. Similar to observations in *P. pastoris* (Rebnegger *et al.* 2016) and several bacteria (Arbige and Chesbro 1982, Chesbro *et al.* 1979, Tappe *et al.* 1999), *O. parapolymorpha* modulated its substrate requirements for cellular maintenance (m_S) in a growth rate-dependent manner. At near-zero growth rates, substrate consumption rates were substantially lower than estimated from faster-growing chemostat cultures. Independent of the specific growth rate and in line with the higher P/O ratio enabled by involvement of Complex I in respiratory oxidation of NADH, the wild-type strain CBS11895 exhibited lower maintenance energy requirements than the Complex I-disrupted strain IMD010 in all glucose-limited chemostat and retentostat cultures. However, Complex I did not play a role in the modulation of m_S at low specific growth rates.

Similar to previous work with *P. pastoris* (Rebnegger *et al.* 2016), analysis of the transcriptome data revealed two regulatory patterns in *O. parapolymorpha*: gene expression that correlated negatively and gene expression that correlated positively with the specific growth rate. In contrast to *P. pastoris*, the majority of these genes were found to correlate positively with the specific growth rate in *O. parapolymorpha* and in

both strain CBS11895 and IMD010 indicated reduced expression levels of biosynthesis-related genes toward lower growth rates. We were not able to relate these growth rate-correlated genes to the mechanisms responsible for the stringent-response-like response in *O. parapolymorpha*. However, because many genes of this yeast lack functional annotation as they are unique to *O. (para)polymorpha* or a small clade of neighboring yeasts, we limited our analysis to the ca. 60% of genes that have orthologs in *S. cerevisiae*, a yeast that does not display a stringent-response-like adaptation at near-zero growth rates (Boender *et al.* 2009, Vos *et al.* 2016). Therefore, mining the subset of genes for which no *S. cerevisiae* ortholog is known could provide novel insights into the mechanism behind the eukaryotic stringent response.

Materials and Methods

Yeast strains, culture conditions, and maintenance. The yeast strains used in this study were CBS11895 (DL-1), a wild-type *Ogataea parapolymorpha* strain ordered from CBS-KNAW (Westerdijk Fungal Biodiversity Institute, Utrecht, The Netherlands), and IMD010, a CBS11895-derived mutant with a disruption in the *NUBM* gene (*nubm*^{G445GC}). Strains were grown in an Innova shaker incubator (New Brunswick Scientific, Edison, NJ, USA) set to 30°C and 200 rpm, in 500-mL shake flasks containing 100 mL of medium. Heat-sterilized (120°C for 20 min) YPD medium (10 g L^{-1} Bacto yeast extract, 20 g L^{-1} Bacto peptone, 20 g L^{-1} glucose, demineralized water) was used for strain construction and maintenance. Solid medium was prepared by addition of 2% (w/v) agar to YPD medium. Frozen stock cultures were prepared from exponentially growing shake-flask cultures by addition of glycerol to a final concentration of 30% (v/v), and aseptically stored in 1-mL aliquots at -80°C.

Molecular biology techniques and strain construction. *Escherichia coli* strain DH5α was used for plasmid transformation, amplification, and storage. Plasmids were isolated from *E. coli* using a GenElute Plasmid Miniprep kit (Sigma-Aldrich, St. Louis, MO, USA). Genomic DNA of yeast colonies used as the template for diagnostic PCR was isolated using the lithium acetate (LiAc)-sodium dodecyl sulfate method (Lõoke *et al.* 2011). Diagnostic PCR was performed using DreamTaq polymerase (Thermo Fisher Scientific, Waltham, MA, USA) and desalted primers (Sigma-Aldrich). DNA fragments obtained by PCR were separated by gel electrophoresis, and PCR purification was performed with a GenElute PCR Clean-Up kit (Sigma-Aldrich).

For the construction of strain IMD010 (*O. parapolymorpha nubm*^{G445GC}), the *O. parapolymorpha* open reading frame (ORF) encoding the Complex I *NUBM* 51-kDa subunit (*OpNUBM,* locus tag HPODL_04625; GenBank accession number XM_014080963.1) was identified. *OpNUBM* was found via a homology search (blastn [https://blast.ncbi.nlm.nih.gov]) in the *O. parapolymorpha* CBS11895 (DL-1) RefSeq assembly (NCBI accession number GCF_000187245.1) (Ravin *et al.* 2013) using the *O. polymorpha* (*Pichia angusta*) partial sequence identified as coding for the Complex I Nubm subunit (GenBank accession number AL434382) (Bridges *et al.* 2009, Souciet *et al.* 2000) as input. Other *O. parapolymorpha* Complex I subunits were assigned based on protein sequence homology (tblastn [https://blast.ncbi.nlm.nih.gov]) with known *P. pastoris* subunits (Bridges *et al.* 2010, Eldarov *et al.* 2011, Ravin *et al.* 2013). *OpNUBM* was disrupted using the pUDP CRISPR/Cas9 system described previously (Juergens *et al.* 2018b). The guideRNA (gRNA) donor plasmid pUD676 was *de novo* synthesized by GeneArt (Thermo Fisher Scientific) and contained the synthetic 233-bp double-stranded DNA (dsDNA) gRNA construct with a gRNA spacer sequence

(5'-CCTGATGTAAATATACGCTG-3') targeting *OpNUBM* after base pair 445 out of 1,467. pUD676 was then integrated into pUDP002 via BsaI-mediated assembly as described previously (Juergens *et al.* 2018b), yielding *OpNUBM*-targeting plasmid pUDP084. For disruption of *OpNUBM,* wild-type *O. parapolymorpha* was transformed with pUDP084 via electroporation, and subjected to a prolonged liquid incubation protocol, as described previously for deletion of *OpADE2* and *OpKU80* (Juergens *et al.* 2018b). Primers 12200 and 12201 (5'-CCCAGCTACGATCTCAAGAC-3' and '5-AACTTGGTGCCCGAGTTAC-3', respectively) were then used for PCR amplification of the *OpNUBM* locus of seven randomly picked single colonies, and subsequent Sanger sequencing (Baseclear, Leiden, The Netherlands) revealed that three out of seven tested colonies contained an indel at the gRNA target site. A single colony of one of the mutants, containing a single cysteine nucleotide insertion between position 445 and 446 of the *OpNUBM* ORF, was restreaked three times subsequently on non-selective YPD medium to remove pUDP084, and renamed IMD010.

Bioreactor cultivation. Bioreactor cultivation was performed using synthetic medium (SM) with the addition of 0.15 g L^{-1} of Pluronic 6100 PE antifoaming agent (BASF, Ludwigshafen, Germany). SM was prepared according to Verduyn *et al.* (1992) and autoclaved at 120°C for 20 min. Glucose and vitamins (Verduyn *et al.* 1992) were prepared separately and filter sterilized (vitamins) or heat sterilized at 110°C for 20 min (glucose). Bioreactors were inoculated with exponentially growing cells from independent shake-flask cultures (grown as described above) with SM and 20 g L^{-1} glucose. All cultures were performed in 2-L benchtop bioreactors (Applikon, Delft, The Netherlands) with initial volumes of 1.4 L (batch) and working volumes of 1.0 or 1.4 L (chemostat) or 1.4 L (retentostat). Cultures were sparged with dried, compressed air (0.5 vvm [volume of gas per volume of liquid per minute]) and stirred at 800 rpm. Temperature was maintained at 30°C, and pH was controlled at 5.0 by automatic addition of a 2 M KOH (batch and chemostat) or 10% (w/v) NH_4OH (retentostat) solution by an ADI 1030 Bio Controller system (Applikon) or by an ez-Control bioreactor controller (Applikon). In chemostats and retentostats, the working volume was maintained by an electrical level sensor that controlled the effluent pump. Culture exhaust gas from bioreactors was cooled with a condenser (2°C) and dried with a Perma Pure dryer (Inacom Instruments, Veenendaal, The Netherlands) prior to online analysis of carbon dioxide and oxygen with a Rosemount NGA 2000 analyzer (Emerson, St. Louis, MO, USA). Batch cultures were performed with 7.5 g L^{-1} of glucose as the sole carbon source and an initial optical density at 660 nm (OD_{660}) of 0.3 (approximately 0.05 g L^{-1} of biomass dry weight). In chemostat cultures, 7.5 g L^{-1} of glucose was used as the sole carbon source, and the dilution rate (D) was set by maintaining a constant inflow rate. Cultures were assumed to have reached a steady state when, after a minimum of

5 volume changes, the oxygen consumption rate, carbon dioxide production rate, and biomass concentration changed by less than 3% over two consecutive volume changes. Retentostat cultivation was performed essentially as described by Rebnegger *et al.* (2016). To predict the accumulation of biomass in retentostat cultures of strains CBS11985 and IMD010, the predictive biomass accumulation script of Rebnegger *et al.* (2016) was used with m_S and Y_{xs}^{max} values as estimated from chemostat cultures with D values of 0.1 and 0.025 h^{-1}. Based on the assumption that maintenance energy requirements are growth rate independent, a feeding regime was selected for *O. parapolymorpha* in which 10 g L^{-1} glucose was used for the preceding chemostat phase, and 5 g L^{-1} was used for the retentostat phase in combination with a 1.2-L mixing vessel, which was identical to the setup used for the earlier work on retentostat cultivation of *P. pastoris* (Rebnegger *et al.* 2016). The dilution rate was determined by maintaining a constant inflow rate of medium from the mixing vessel, a 3-L benchtop bioreactor (Applikon) with a working volume of 1.2 L and stirred at 500 rpm. The volume in the mixing vessel was kept constant by an electrical level sensor that controlled the feed pump of the mixing vessel. The medium was SM as described above but contained an additional 0.5 mL L^{-1} of the concentrated trace element solution (1.5X final concentration) and an additional 1 mL L^{-1} of the vitamin stock solution (2X final concentration) (Verduyn *et al.* 1992). Retentostat cultivation was preceded by a chemostat cultivation (D of 0.025 h^{-1}) using the same conditions as described for the subsequent retentostat cultivation. During the chemostat phase, the medium flowing into the mixing vessel contained 10 g L^{-1} glucose. Once a steady state was achieved, the retentostat phase was initiated by two changes: (i) the medium flowing into the mixing vessel was changed to be drawn from a medium vessel with identical medium composition but with 5 g L^{-1} glucose as the limiting compound, and (ii) the culture effluent was redirected through a filtered effluent port, equipped with a hollow stainless steel filter support with an autoclavable hydrophobic polypropylene filter with 0.22-µm pore size (Trace Analytics, Braunschweig, Germany). Prior to heat sterilization, the filter was soaked overnight in a 96% ethanol solution and subsequently rinsed with 1X phosphate-buffered saline (Sigma Aldrich).

Biomass measurements. Optical density was measured at 660 nm on a Jenway 7200 spectrophotometer (Jenway, Staffordshire, UK). For biomass dry weight determination (typically performed in triplicate), exactly 10 mL of culture broth was filtered over predried and preweighed membrane filters (0.45-µm pore size; Pall Corporation, Ann Arbor, MI, USA), which were washed with demineralized water, dried in a microwave oven at 350 W for 20 min, and weighed immediately (Postma *et al.* 1989). Samples from chemostat and retentostat cultures were diluted with demineralized water prior to filtration to obtain a biomass dry weight concentration of approximately 2 g L^{-1}. The

exact dilution was calculated by weighing the amount of sample and diluent and assuming a density of 1 g mL^{-1} for both fractions. For samples from late-stage retentostat cultures of strain IMD010 (after approximately 14 days and onwards), membrane filters were placed in glass bowls and covered with plastic funnels for microwave drying as under normal drying conditions biomass flakes formed that detached from the membrane filters, preventing accurate determination. Membrane filters were routinely re-dried and reweighed to ensure complete drying. Biomass protein content was determined using dried bovine serum albumin (BSA) (fatty acid free; Sigma-Aldrich) as described previously (Verduyn *et al.* 1990), with the modifications that NaOH was used instead of KOH and absorbance was measured at 510 instead of at 550 nm. Culture samples were diluted with demineralized water to biomass dry weight concentrations between 2.2 and 3.8 g L^{-1} prior to protein content analysis.

Metabolite analysis. For the determination of extracellular metabolite concentrations during batch fermentations, 1-mL aliquots of culture sample were centrifuged for 3 min at 20,000 g and the supernatant was used for analysis. Samples from chemostat and retentostat cultures were rapidly quenched with the cold steel beads method (Mashego *et al.* 2003). Metabolite concentrations were analyzed by high-performance liquid chromatography (HPLC) on an Agilent 1100 HPLC (Agilent Technologies, Santa Clara, CA, USA) with an Aminex HPX-87H ion exchange column (Bio-Rad, Veenendaal, The Netherlands) operated at 60°C with 5 mM H_2SO_4 as the mobile phase at a flow rate of 0.6 mL min^{-1}.

Viability assays. For cell viability determination based on membrane integrity (via propidium iodide [PI]), approximately 0.5 mL of culture broth was sampled into 15 mL of ice-cold 10 mM Na-HEPES buffer (pH 7.2) containing 2% (w/v) glucose and kept on ice. Cell concentrations were determined using a Z2 Coulter counter (Beckman Coulter, Fullerton, CA, USA) set to a detection interval of 1.5 to 5.8 µm. The buffered sample was then diluted in isotone II diluent (Beckman Coulter, Woerden, The Netherlands) to a suspension containing 10^7 cells mL^{-1} and stained with PI (Sigma-Aldrich) as described previously (Boender *et al.* 2011). The stained samples were analyzed on an Accuri C6 flow cytometer (BD Biosciences, Franklin Lakes, NJ, USA), equipped with a 488-nm laser, and detected by the FL-3 channel (620-nm band pass filter) for PI staining. Per sample, 30,000 events (cells) were analyzed. The viability was determined using Flowing software, version 2.5.1 (Perrtu Terho, Turku Centre for Biotechnology, University of Turku, Finland), by subtracting the percentage of PI-stained cells from a starting value of 100%. For determination of cell viability based on CFU counts, cultures were sampled into Na-HEPES buffer as described above and analyzed on a BD FACSAria

II SORP cell sorter (BD Biosciences), equipped with a 70-µm nozzle and operated with filtered FACSFlow sheath fluid (BD Biosciences). Evaluation of cytometer performance, analysis of cell morphology, and cell sorting were essentially performed as described previously (Brickwedde *et al.* 2018). Gating of cell populations for CFU count determination by plating was performed so that typically more than 90% of all detected events (cells) would be sorted. Viability was determined as the average percentage of sorted cells able to form a colony after 3 days of incubation at 30°C on quintuplicate YPD plates (96 cells sorted per plate).

Calculation of growth rate dependency of maintenance energy requirements. For chemostat cultures, specific growth rate (µ) and biomass-specific glucose uptake rate (q_S) were calculated by solving biomass and substrate mass balances assuming steady-state conditions, and least squares linear regression was used to estimate maintenance energy substrate requirements (m_S; intercept with y-axis) and theoretical maximum biomass yield (Y_{XS}^{max}; reciprocal of slope) coefficients (Pirt 1965) from q_S/μ relationships. Calculations for retentostat cultures were performed essentially as described by Rebnegger *et al.* (2016): pairs of µ and q_S values were calculated from biomass accumulation between adjacent sampling points, and m_S values were estimated via least squares linear regression from moving windows of continuous pairs of calculated µ and q_S values (including from chemostat cultivations), with the exception that moving windows of 5 q_S-µ pairs were used for m_S estimation.

Whole-genome sequencing and stability of *NUBM* disruption. Genomic DNA of CBS11895 and IMD010 was isolated using a Qiagen 100/G kit (Qiagen, Hilden, Germany) from a shake-flask culture grown in YPD medium to stationary phase, according to the manufacturer's instructions. DNA concentrations were quantified using a Qubit fluorometer, version 2.0 (Thermo Fisher Scientific). CBS11895 was sequenced by Novogene Bioinformatics Technology Co., Ltd. (Yuen Long, Hong Kong) on a HiSeq 2500 instrument (Illumina, San Diego, CA, USA) with 150-bp paired-end reads using a True-seq PCR-free library preparation (Illumina). IMD010 was sequenced on a MiSeq instrument (Illumina) using a TruSeq DNA PCR-free library preparation as described previously (Swiat *et al.* 2017).

In order to verify the genetic stability of the *nubm*^{G445GC} disruption in strain IMD010 during the prolonged glucose-limited cultivations, a minimum of four single-colony isolates from each individual chemostat and retentostat cultivation with IMD010 was tested for the presence of the mutation. To this end, the cultures were plated for single colonies on solid YPD medium at the last sampling point of each fermentation, their genomic DNA was isolated, and primers 12200 and 12201 were used to PCR amplify and Sanger sequence the site containing the *nubm*^{G445GC} disruption, as described

above. The *nubm*^{G445GC} genotype was still present in all investigated colonies, and no additional mutations were detected within any of the sequenced 688-bp amplicons.

RNA extraction, RNA sequencing, and transcriptome data analysis. Sampling for transcriptome analysis was performed by quenching culture broth directly into liquid nitrogen to immediately stop mRNA turnover (Piper *et al.* 2002), followed by storage at -80°C. In the case of batch cultures, sampling for transcriptome analysis was done in mid-exponential phase at a biomass dry weight concentration of approximately 0.9 g L^{-1} with >75% of the initial glucose concentration remaining in the reactor. Processing of samples for long-term storage using AE buffer, acid-phenol-chloroform-isoamyl alcohol (125:24:1, pH 4.5; Thermo Fisher Scientific), and sodium dodecyl sulfate, as well as total RNA isolation was performed as described previously (Brickwedde *et al.* 2018). The quality of the total extracted RNA was evaluated with an Agilent 2200 Tapestation (Agilent Technologies, Santa Clara, CA), and the RNA concentration was determined using a Qubit 2.0 fluorometer (Thermo Fisher Scientific) combined with a Qubit RNA BR (broad-range) assay kit (Thermo Fisher Scientific). Library preparation and RNA sequencing were performed by Novogene Bioinformatics Technology Co., Ltd. (Yuen Long, Hong Kong). Sequencing was done with an Illumina paired-end 150-bp sequencing read system (PE150) using a 250- to 300-bp insert strand-specific library which was prepared by Novogene. For the library preparation, mRNA enrichment was done using oligo(dT) beads. After random fragmentation of the mRNA, cDNA was synthetized from the mRNA using random hexamer primers. Afterwards, second-strand synthesis was done by addition of a custom second-strand synthesis buffer (Illumina), deoxynucleoside triphosphates (dNTPs), RNase H, and DNA polymerase I. Finally, after terminal repair, A ligation, and adaptor ligation, the double-stranded cDNA library was finalized by size selection and PCR enrichment.

The sequencing data for the samples obtained by Novogene had an average read depth of 21 million reads per sample. For each sample, reads were aligned to the genome of CBS11895 (DL-1) RefSeq assembly (NCBI accession number GCF_000187245.1) (Ravin *et al.* 2013) with the two-pass STAR procedure (Dobin *et al.* 2013). In the first pass, a splice junction database was assembled which was used to inform the second round of alignments. Introns were allowed to be between 15 and 4,000 bp, and soft clipping was disabled to prevent low-quality reads from being spuriously aligned. Ambiguously mapped reads were removed. Expression was quantified per transcript using HTSeq count in union intersection mode (Anders *et al.* 2015). To exclude from the analysis genes expressed at low levels, genes with an average fragments per kilobase per million (FPKM) value below 10 in all samples were removed. Counts were normalized by TMM normalization using the edgeR package (Robinson *et al.* 2010), and subsequently differentially expressed genes were

determined with an absolute $\log_2$ fold change of >2 and a false-discovery rate of <0.001. Mean normalization of transcript data was performed per gene and separately for strains CBS11895 and IMD010 using the TMM-normalized FPKM values and was done either including (for comparative expression analysis) or excluding (for analysis of growth rate-correlated gene expression) data from the batch fermentations. For identification of growth rate-correlated gene clusters, analysis of the mean-normalized transcript values versus specific growth rate was performed using the maSigPro R-package (Conesa *et al.* 2006, Nueda *et al.* 2014). Genes with a trend significantly different from the mean were selected with a Benjamini-Hochberg corrected P value of <0.1, and subsequently the regression parameters for two clusters of genes were identified with a significance cutoff of 0.05 and an R^2 of >0.8. For determination of enriched Gene Ontology (GO) terms in growth rate-correlated gene clusters and sets of differentially expressed genes, the online generic GO Term Finder tool (http://go.princeton.edu/cgi-bin/GOTermFinder) and *Saccharomyces* Genome Database annotation were used. A cutoff of 0.01 was used for the corrected P value (Bonferroni correction), and a background list was provided containing all *O. parapolymorpha* CBS11895 protein-coding genes for which *S. cerevisiae* S288C orthologs could be identified (3,094 out of 5,325; obtained using the Orthologous Matrix Database (Altenhoff *et al.* 2018)). The number of GO terms was reduced using REVIGO with an allowed similarity setting of 0.5 (Supek *et al.* 2011) (see Data Set S1 in the supplementary material for all identified GO terms).

Proteome processing and data analysis. For proteome sampling, a culture sample equivalent to 2 to 4 mg of biomass dry weight was sampled into precooled microcentrifuge tubes, pelleted by centrifugation at 4°C at 4700 g for 5 min, and washed with 1.5 mL of ice-cold 1X phosphate-buffered saline (Sigma-Aldrich). After an additional centrifugation step under identical conditions and subsequent removal of the phosphate buffered saline, cell pellets were stored at -80°C until further processing. In the case of batch cultures, sampling for proteome analysis was done in mid-exponential phase at a biomass dry weight concentration of approximately 0.9 g L^{-1} with >75% of the initial glucose concentration remaining in the reactor. To process samples for analysis, cell mass was normalized to a dry weight of 1.6 mg and mechanically lysed using 0.5-mm zirconium beads and a PreCellys homogenizer. Proteins were isolated using Bligh and Dyer extraction (Sapcariu *et al.* 2014), followed by reduction, alkylation, and digestion using trypsin. Samples were analyzed in technical triplicates by liquid chromatography tandem mass spectrometry (LC-MS/MS) using a Vanquish UHPLC coupled to a Q Exactive Plus Orbitrap MS (Thermo Fisher Scientific). Peptides were separated using reverse-phase chromatography using a gradient of water with 0.1% formic acid (solvent A) and acetonitrile with 0.1% formic acid (solvent B) from 2% B to

45% B in 50 min. Data-dependent acquisition (DDA) was performed with a resolution setting at 70,000 within the 400- to 1,600-m/z range and a maximum injection time of 75 ms, followed by high-energy collision-induced dissociation-activated (HCD) MS/MS on the top 15 most abundant precursors using a resolution setting of 17,500 and a 200- to 2,000-m/z range with a maximum injection time of 50 ms. The minimum intensity threshold for MS/MS was 1,000 counts, and peptide species with 1 and >8 charges were excluded. MS/MS spectra were analyzed with the SEQUEST HT search engine and Proteome Discoverer, version 2.3, against the proteins of the CBS11895 (DL-1) RefSeq assembly (NCBI accession number GCF_000187245.1) (Ravin *et al.* 2013). Label-free quantification was performed using the top three unique peptides measured for each protein. Retention time alignment was performed on the most abundant signals obtained from nonmodified peptides measured in all samples, and results were corrected for the total ion intensities measured for each sample. For subsequent analysis, only proteins were taken along that achieved a combined detection confidence with an FDR of <1% and additionally were individually detected with an FDR of <1% in at least 5 out of the total 48 LC-MS analyses (6 per condition). For proteins that passed these requirements, protein abundance was set to 0 for individual analyses that did not exhibit a detection confidence with an FDR of <1%, and the average abundance of all analyses per condition was used for further calculation. Proteins were considered not detected for a specific condition if they were not measured at least once with a detection confidence of an FDR of <1% for that condition. Mean normalization of the protein data was performed per gene and separately for strains CBS11895 and IMD010 using the total ion intensity-normalized protein abundances.

Data availability. Transcript abundances, lists of differentially expressed genes, sets of growth rate-correlated genes, identified *S. cerevisiae* orthologs of *O. parapolymorpha* protein-coding genes, complete lists of enriched GO terms, and total ion intensity-normalized protein abundances are available in Data Set S1 in the supplementary material.

Genome sequencing data of CBS11895 and IMD010 are available at NCBI (https://www.ncbi.nlm.nih.gov/) under BioProject accession number PRJNA588376. RNA-seq data are available at NCBI (https://www.ncbi.nlm.nih.gov/) under Gene Expression Omnibus (GEO) accession number GSE140480. Raw proteomics data are available on figshare (https://doi.org/10.6084/m9.figshare.11398773) (Juergens *et al.* 2020b).

Acknowledgements

We thank Pilar de la Torre Cortéz for whole-genome sequencing of strain IMD010, Marcel van den Broek for help with genome/transcriptome analysis and identification of orthologous genes, and Janine Nijenhuis for her contribution to Figure 4.1.

Supplementary material

The full supplementary material including Data Set S1 are available at the AEM journal, https://doi.org/10.1128/AEM.00678-20.

Table S4.1: Physiology of wild type strain *O. parapolymorpha* CBS11895 and Complex I-disrupted mutant IMD010 in aerobic glucose-limited chemostat cultures at a dilution rate of 0.025 h^{-1}. Reported values are mean ± SEM calculated from two independent steady state cultures. Reported *p*-values (Student's *t*-test) refer to the difference between mean values observed for CBS11895 and IMD010. Carbon recovery calculations are based on a biomass carbon content of 48% (w/w). BDL: below detection limit (10 μM). Reported cell viability based on propidium iodide staining. Symbols: $Y_{X/S}$ and $Y_{X/O2}$ = yield of biomass dry weight on glucose and oxygen, respectively; RQ = respiratory quotient; $q_{Glucose}$, q_{CO2} and q_{O2} = biomass-specific uptake/production rate of glucose, CO_2 and O_2, respectively; C_X = biomass dry weight concentration. The presented data was determined from chemostat cultures preceding retentostat cultivations, with exception of the biomass protein content, which was determined from separate, duplicate chemostat cultures.

Strain	**CBS11895**	**IMD010**	***P*-value**
Actual dilution rate (h^{-1})	0.025 ± 0.000	0.025 ± 0.000	-
Reservoir glucose (g L^{-1})	10.03 ± 0.01	9.75 ± 0.05	-
Residual glucose (mM)	BDL	BDL	-
$Y_{X/S}$ (g biomass [g glucose]$^{-1}$)	0.42 ± 0.00	0.33 ± 0.00	<.001
$Y_{X/O2}$ (g biomass [g O_2]$^{-1}$)	0.78 ± 0.03	0.52 ± 0.01	.030
RQ	0.97 ± 0.02	1.00 ± 0.03	.38
$q_{Glucose}$ (mmol [g biomass]$^{-1}$ h^{-1})	-0.34 ± 0.00	-0.42 ± 0.01	.044
q_{CO2} (mmol [g biomass]$^{-1}$ h^{-1})	0.98 ± 0.01	1.53 ± 0.02	.011
q_{O2} (mmol [g biomass]$^{-1}$ h^{-1})	-1.02 ± 0.03	-1.52 ± 0.06	.033
C_X (g biomass L^{-1})	4.20 ± 0.01	3.24 ± 0.00	.007
Protein content (g protein [g biomass]$^{-1}$)	0.36 ± 0.01	0.39 ± 0.02	.26
Cell viability (%)	99.4 ± 0.2	99.8 ± 0.1	.20
Carbon recovery (%)	99.0 ± 0.7	100.0 ± 0.7	.41

Table S4.2: Enriched GO terms (p value cutoff 0.01; Bonferroni correction) of categories 'biological process' (BP), 'molecular function' (MF) and 'cellular component' (CC) found in set of significantly upregulated genes in *O. parapolymorpha* strain IMD010 (devoid of functional Complex I) compared to CBS11895 (wild type) in aerobic glucose-grown batch cultures. GO enrichment analysis was based on 275 (out of 409) genes for which an *S. cerevisiae* ortholog could be identified, and a background (genome) of 3095 genes was used for testing. Redundant GO terms were removed using REVIGO. See data appendix for full list of GO terms.

GO term	Type	Description	P value (corrected)	Genes in set	Genes in genome
GO:0022613	BP	ribonucleoprotein complex biogenesis	5.99E-18	89	381
GO:0010467	BP	gene expression	6.24E-17	162	1033
GO:0044085	BP	cellular component biogenesis	1.42E-15	135	796
GO:0006364	BP	rRNA processing	3.40E-12	57	219
GO:0002181	BP	cytoplasmic translation	2.18E-09	37	121
GO:0034641	BP	cellular nitrogen compound metabolic process	1.02E-08	177	1395
GO:0006396	BP	RNA processing	6.60E-08	73	397
GO:0043603	BP	cellular amide metabolic process	3.77E-06	70	406
GO:0034660	BP	ncRNA metabolic process	4.82E-06	67	383
GO:0071840	BP	cellular component organization or biogenesis	1.50E-05	165	1363
GO:0043170	BP	macromolecule metabolic process	2.57E-04	181	1592
GO:0043933	BP	macromolecular complex subunit organization	5.45E-04	72	474
GO:0044271	BP	cellular nitrogen compound biosynthetic process	2.49E-03	113	891
GO:0033108	BP	mitochondrial respiratory chain complex assembly	4.62E-03	12	31
GO:0003735	MF	structural constituent of ribosome	2.23E-23	58	146
GO:0005198	MF	structural molecule activity	8.80E-18	62	205
GO:0003723	MF	RNA binding	1.09E-04	67	427
GO:0019843	MF	rRNA binding	4.98E-04	17	55
GO:0003676	MF	nucleic acid binding	5.71E-03	88	677
GO:1990904	CC	ribonucleoprotein complex	1.03E-28	116	462
GO:0044391	CC	ribosomal subunit	1.81E-21	58	156
GO:0043228	CC	non-membrane-bounded organelle	1.03E-14	138	850
GO:0032991	CC	macromolecular complex	1.83E-13	186	1381
GO:0031974	CC	membrane-enclosed lumen	2.60E-12	125	776
GO:0005730	CC	Nucleolus	3.98E-12	56	218
GO:0098798	CC	mitochondrial protein complex	1.05E-08	41	156
GO:0044445	CC	cytosolic part	1.89E-06	36	148
GO:0044428	CC	nuclear part	5.91E-06	99	683
GO:0044446	CC	intracellular organelle part	1.38E-05	205	1848
GO:0005688	CC	U6 snRNP	1.86E-05	6	6
GO:0005732	CC	small nucleolar ribonucleoprotein complex	1.27E-04	9	19
GO:1990726	CC	Lsm1-7-Pat1 complex	3.44E-03	5	6

Table S4.3: Enriched GO terms (p value cutoff 0.01; Bonferroni correction) of categories 'biological process' (BP), 'molecular function' (MF) and 'cellular component' (CC) found in sets of genes that exhibited positive or negative correlation of expression with specific growth rate in *O. parapolymorpha* CBS11895 (wild type) and/or IMD010 (devoid of functional Complex I), based on samples taken from glucose-limited chemostat (0.1 and 0.025 h^{-1}) and late-stage retentostat (0.001 h^{-1}) cultures. Test sets: a (positive correlation only in CBS11895; 459 genes), b (positive correlation in both CBS11895 and IMD010; 414 genes), c (positive correlation only in IMD010; 169 genes), d (negative correlation only in CBS11895; 137 genes). All gene sets were tested against a background (genome) of 3095 genes. Redundant GO terms were removed using REVIGO. See data appendix for full list of GO terms.

Set	GO term	Description	Type	P value (corrected)	Genes in set	Genes in genome
a	GO:0051276	chromosome organization	BP	1.14E-05	84	317
a	GO:0007059	chromosome segregation	BP	1.33E-04	35	96
a	GO:0098813	nuclear chromosome segregation	BP	1.12E-03	29	78
a	GO:0000280	nuclear division	BP	3.49E-03	35	108
a	GO:0071824	protein-DNA complex subunit organization	BP	5.62E-03	35	110
a	GO:0097159	organic cyclic compound binding	MF	1.94E-03	216	1164
a	GO:0003677	DNA binding	MF	4.32E-03	65	266
a	GO:1901363	heterocyclic compound binding	MF	5.10E-03	213	1158
a	GO:0005488	binding	MF	5.79E-03	316	1857
a	GO:0015630	microtubule cytoskeleton	CC	4.00E-06	27	59
a	GO:0043228	non-membrane-bounded organelle	CC	5.38E-04	169	850
a	GO:0000794	condensed nuclear chromosome	CC	8.34E-04	19	42
a	GO:0005856	cytoskeleton	CC	1.32E-03	39	128
a	GO:0044428	nuclear part	CC	3.66E-03	138	683
a	GO:0005634	nucleus	CC	7.87E-03	229	1269
b	GO:0009058	biosynthetic process	BP	1.29E-07	231	1274
b	GO:1901576	organic substance biosynthetic process	BP	1.32E-07	230	1267
b	GO:0044281	small molecule metabolic process	BP	8.22E-07	112	558
b	GO:0006520	cellular amino acid metabolic process	BP	9.12E-05	54	198
b	GO:0017144	drug metabolic process	BP	3.21E-03	48	186
b	GO:0003824	catalytic activity	MF	6.24E-05	263	1602
b	GO:0071162	CMG complex	CC	4.19E-03	8	11
c	GO:0140101	catalytic activity, acting on a tRNA	MF	2.14E-03	15	76
d	GO:0016021	integral component of membrane	CC	1.87E-10	64	632
d	GO:0005783	endoplasmic reticulum	CC	1.55E-08	47	413
d	GO:0012505	endomembrane system	CC	4.65E-07	60	675
d	GO:0016020	membrane	CC	3.03E-06	81	1121
d	GO:0044444	cytoplasmic part	CC	3.54E-05	115	1999
d	GO:0031984	organelle subcompartment	CC	3.56E-05	37	356
d	GO:0098827	endoplasmic reticulum subcompartment	CC	2.14E-04	29	259
d	GO:0005737	cytoplasm	CC	2.16E-03	129	2529

Table S4.4: Protein abundances of nuclearly encoded Complex I subunits and alternative NADH dehydrogenases (NDH2) in *O. parapolymorpha* strains CBS11895 (wild type) and IMD010 (disrupted Complex I Nubm subunit). Samples were taken from aerobic glucose-grown batch (0.37 h^{-1}), chemostat (0.1 and 0.025 h^{-1}) and retentostat (0.001 h^{-1}) cultures. Data is presented as mean ± standard deviation of two independent biological replicate analyses (each based on three technical replicates). The number between parentheses denotes how many times (out of six technical replicates) a specific protein was detected with a detection confidence of FDR <1%. N.D.: 'not detected' (see Methods section). The Complex I subunits were assigned based on sequence homology with subunits from *Pichia pastoris* (Bridges *et al.* 2010). For *P. pastoris* Complex I subunits NB5M, NB8M and NUTM, corresponding ORFs can be detected in the genome of *O. parapolymorpha*, but were not annotated as protein-coding in the *O. parapolymorpha* reference assembly (Ravin *et al.* 2013) and excluded from analysis. Complex I subunit NUUM (HPODL_05121) was excluded from proteome analysis because abundance data did not pass quality requirements for analysis (see Methods section).

Subunit	Locus tag	Type	CBS11895				IMD010			
			0.37 h^{-1}	0.1h^{-1}	0.025 h^{-1}	0.001 h^{-1}	0.37 h^{-1}	0.1h^{-1}	0.025 h^{-1}	0.001 h^{-1}
NUAM	HPODL_03689	Essential	179824 ± 38% (6)	481529 ± 0% (6)	442908 ± 25% (6)	322146 ± 30% (6)	96960 ± 12% (5)	266627 ± 9% (6)	30351 ± 7% (4)	N.D.
NUBM	HPODL_04625	Essential	71482 ± 56% (6)	280305 ± 2% (6)	316559 ± 5% (6)	293642 ± 31% (6)	N.D.	N.D.	N.D.	N.D.
NUCM	HPODL_01297	Essential	24688 (1)	375650 ± 2% (6)	344104 ± 10% (6)	144700 ± 67% (4)	N.D.	353512 ± 5% (6)	211590 ± 1% (6)	N.D.
NUGM	HPODL_03393	Essential	50286 (2)	399470 ± 0% (6)	370648 ± 4% (6)	203294 ± 39% (5)	15278 (1)	337246 ± 2% (6)	93404 ± 67% (3)	N.D.
NUHM	HPODL_01287	Essential	32053 ± 41% (2)	394675 ± 3% (6)	384382 ± 13% (6)	242639 ± 14% (6)	N.D.	N.D.	N.D.	N.D.
NUIM	HPODL_02101	Essential	N.D.	77726 ± 54% (3)	148803 ± 16% (6)	14726 (1)	N.D.	N.D.	20701 ± 11% (2)	N.D.
NUKM	HPODL_02758	Essential	8047 (1)	206143 ± 4% (6)	202897 ± 15% (6)	34087 ± 22% (2)	N.D.	106113 ± 1% (6)	23979 ± 28% (2)	N.D.
NUEM	HPODL_03913	Accessory	25354 ± 47% (2)	359541 ± 15% (6)	382660 ± 16% (6)	277320 ± 3% (6)	47377 ± 35% (2)	254346 ± 1% (6)	191010 ± 7% (6)	N.D.
NESM	HPODL_04326	Accessory	N.D.	329988 ± 4% (6)	285892 ± 6% (6)	29033 (1)	N.D.	354216 ± 5% (6)	226425 ± 4% (6)	N.D.
NUJM	HPODL_01263	Accessory	N.D.	251225 ± 11% (6)	257097 ± 9% (6)	126511 ± 25% (5)	N.D.	238167 ± 11% (6)	160546 ± 29% (6)	N.D.
NUXM	HPODL_00687	Accessory	N.D.	79080 (2)	86825 ± 16% (2)	15807 (1)	N.D.	113479 ± 65% (3)	33902 (1)	N.D.
NUPM	HPODL_04116	Accessory	42052 ± 28% (2)	235964 ± 28% (5)	218736 ± 10% (5)	29678 (1)	47353 ± 24% (2)	336688 ± 3% (6)	170028 ± 22% (5)	N.D.
NUZM	HPODL_02157	Accessory	23774 (1)	405364 ± 0% (6)	432082 ± 8% (6)	313045 ± 14% (6)	N.D.	282122 ± 4% (6)	138582 ± 84% (4)	N.D.
NUSM	HPODL_00962	Accessory	5151 (1)	357940 ± 25% (5)	416528 ± 9% (6)	166118 ± 85% (4)	N.D.	462020 ± 3% (6)	135084 ± 61% (3)	5393 (1)
NIMM	HPODL_04436	Accessory	N.D.	184355 ± 31% (5)	138416 ± 81% (4)	21623 (1)	N.D.	206603 ± 29% (5)	74854 ± 20% (3)	N.D.
NB6M	HPODL_04286	Accessory	N.D.	95476 ± 42% (3)	77662 ± 34% (3)	22259 (1)	N.D.	94476 ± 40% (3)	24656 (1)	N.D.
N7BM	HPODL_02876	Accessory	N.D.	207768 ± 25% (5)	236554 ± 0% (6)	21041 (1)	N.D.	N.D.	N.D.	N.D.
NUYM	HPODL_02906	Accessory	N.D.	284591 ± 5% (6)	273960 ± 11% (6)	175143 ± 49% (5)	N.D.	N.D.	N.D.	N.D.
NIAM	HPODL_04805	Accessory	30291 ± 29% (2)	279664 ± 0% (6)	150171 ± 45% (5)	19912 (1)	28892 ± 20% (2)	292267 ± 0% (6)	88641 ± 48% (3)	N.D.
NUMM	HPODL_01160	Accessory	N.D.	146881 ± 10% (6)	136240 ± 1% (6)	37833 ± 44% (3)	3085 (1)	16775 (2)	4790 (1)	N.D.
NUFM	HPODL_00771	Accessory	N.D.	242389 ± 14% (6)	302979 ± 2% (6)	181680 ± 18% (5)	N.D.	204788 ± 1% (6)	81068 ± 26% (3)	N.D.
NB4M	HPODL_02851	Accessory	N.D.	145115 ± 6% (6)	158945 ± 12% (6)	15871 (1)	N.D.	16112 (1)	25179 ± 7% (2)	N.D.
NI2M	HPODL_05260	Accessory	62507 (2)	346886 ± 1% (6)	387578 ± 17% (6)	237959 ± 15% (5)	42787 ± 44% (2)	417024 ± 17% (6)	245943 ± 4% (6)	N.D.
NIPM	HPODL_02838	Accessory	N.D.	182097 ± 72% (4)	125963 ± 52% (3)	24355 (1)	N.D.	116294 ± 67% (3)	84215 ± 52% (3)	N.D.
ACPM2	HPODL_02246	Accessory	N.D.	540289 ± 5% (6)	499646 ± 11% (6)	161965 ± 82% (3)	N.D.	517691 ± 16% (6)	381346 ± 2% (6)	259593 ± 38% (5)
NIDM	HPODL_04828	Accessory	N.D.	363610 ± 4% (6)	327785 ± 4% (6)	6049 (1)	9165 ± 6% (2)	397009 ± 18% (6)	196454 ± 33% (5)	N.D.
NI8M	HPODL_04214	Accessory	N.D.	343986 ± 2% (6)	298730 ± 10% (6)	28227 (1)	32510 ± 0% (2)	200145 ± 19% (6)	27715 ± 18% (2)	N.D.
ACPM1	HPODL_01821	Accessory	618989 ± 9% (6)	883525 ± 6% (6)	761495 ± 1% (6)	581073 ± 23% (6)	719983 ± 4% (6)	815408 ± 18% (6)	734720 ± 2% (6)	754365 ± 6% (6)
NI9M	HPODL_05148	Accessory	23012 (1)	287906 ± 23% (5)	292372 ± 8% (6)	87698 ± 76% (3)	41804 ± 26% (2)	335671 ± 25% (6)	168099 ± 15% (4)	12915 (1)
NDH2-1	HPODL_02792	NDH2	274434 ± 37% (6)	N.D.	N.D.	N.D.	203027 ± 36% (5)	55633 ± 60% (3)	50900 ± 83% (3)	N.D.
NDH2-2	HPODL_00256	NDH2	540086 ± 0% (6)	337718 ± 11% (6)	490208 ± 21% (6)	467389 ± 4% (6)	517421 ± 5% (6)	295604 ± 5% (6)	394630 ± 0% (6)	450323 ± 7% (6)
NDH2-3	HPODL_02018	NDH2	118201 ± 2% (6)	N.D.	20315 ± 87% (3)	44509 ± 34% (4)	112193 ± 3% (6)	N.D.	7744 (1)	50213 ± 31% (5)

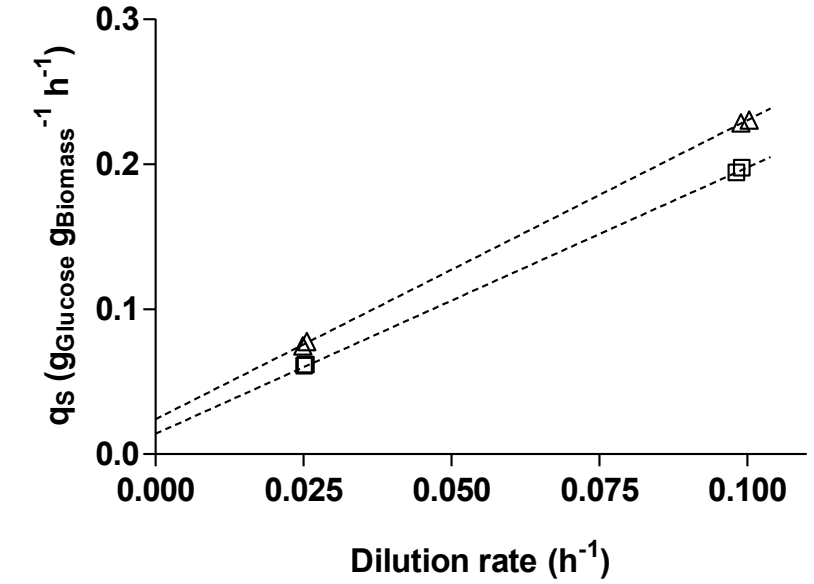


Figure S4.1: Biomass-specific glucose uptake rate (q_S) of wild type *O. parapolymorpha* CBS11895 (squares) and Complex I-deficient strain IMD010 (triangles) in aerobic, glucose-limited chemostat cultures. Each data point represents an independent biological replicate. Least squares linear regression (dashed lines) was used to estimate maintenance energy requirements (m_S; intercept with y axis) and theoretical maximum biomass yield ($Y_{X/S}^{max}$; reciprocal of slope) coefficients.

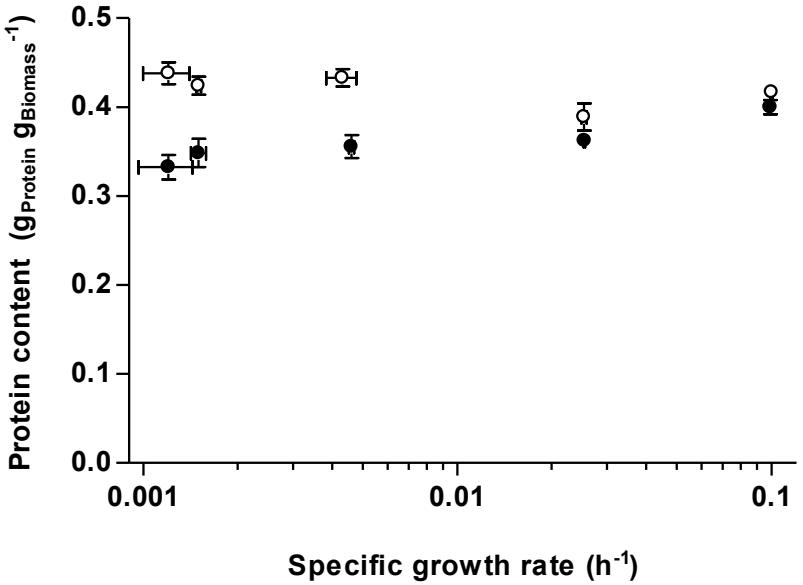


Figure S4.3: Biomass protein content of *O. parapolymorpha* strains CBS11895 (wild type; closed circles) and IMD010 (Complex I-deficient; open circles) at different specific growth rates. Samples were taken from steady state aerobic glucose-limited chemostat cultures (specific growth rates of 0.1 h^{-1} and 0.025 h^{-1}) and from aerobic glucose-limited retentostat cultures after 4, 11 and 21 days of cultivation (specific growth rates below 0.025 h^{-1}). For each data point samples were taken from two independent cultures, and data is presented as mean ± standard deviation. Horizontal error bars represent standard deviation of specific growth rate determined for retentostat cultures. For retentostat culture samples, the mean protein content of CBS11895 was found to be significantly lower than that of IMD010 at each equivalent sampling point (Student's *t*-test, $p < .05$).

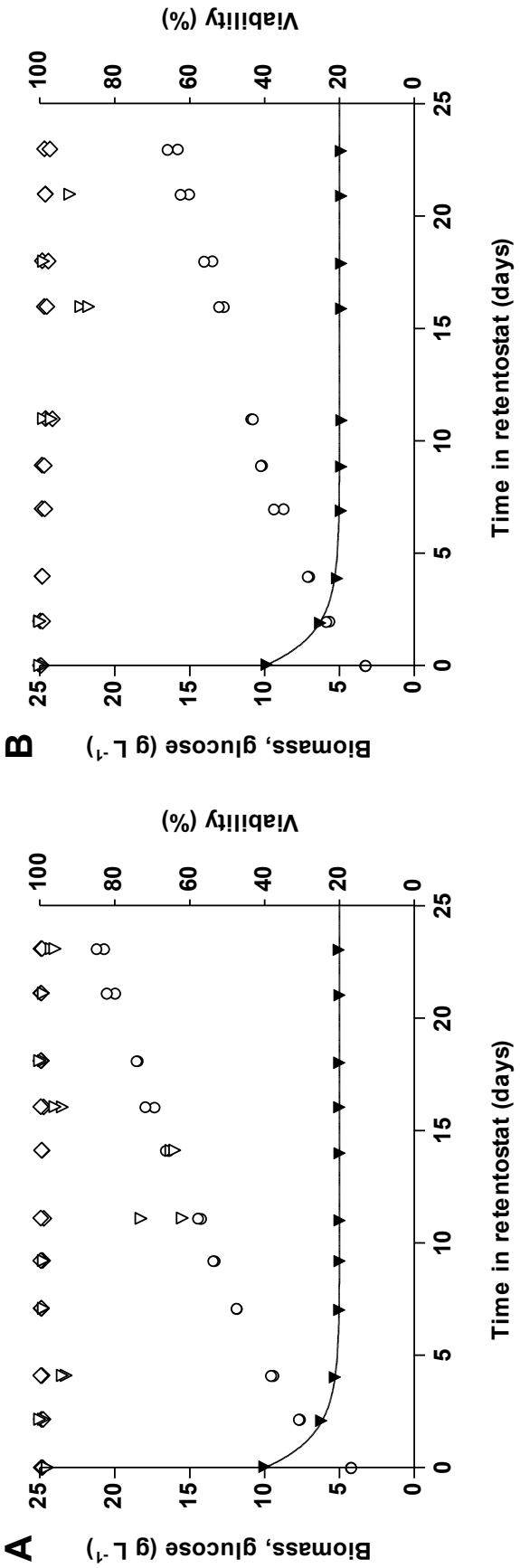


Figure S4.2: Biomass accumulation profile, viability and feed glucose concentration of aerobic glucose-limited *O. parapolymorpha* retentostat cultures of wild-type strain CBS11895 (A) and the congenic Complex I-deficient mutant IMD010 (B). Depicted are the measured biomass dry weight concentration (open circles), culture viability based on propidium iodide staining (open diamonds) and CFU determination (open triangles), and predicted (solid line) and measured (closed triangles) feed glucose concentration in the mixing vessel of two independent cultures from each strain.

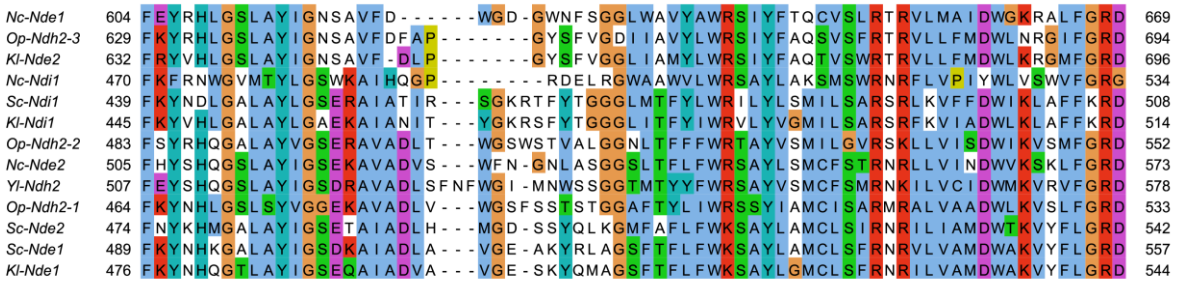


Figure S4.4: Section of MUSCLE sequence alignment showing the characteristic conserved C-terminal domain (Feng *et al.* 2012) of the three putative *O. parapolymorpha* alternative NAD(P)H dehydrogenases and known fungal/yeast alternative NAD(P)H dehydrogenases. Visualized by Jalview using Clustalx residue colors. *Kl, Kluyveromyces lactis, Nc, Neurospora crassa, Op, Ogataea parapolymorpha, Sc, Saccharomyces cerevisiae, Yl, Yarrowia lipolytica.*

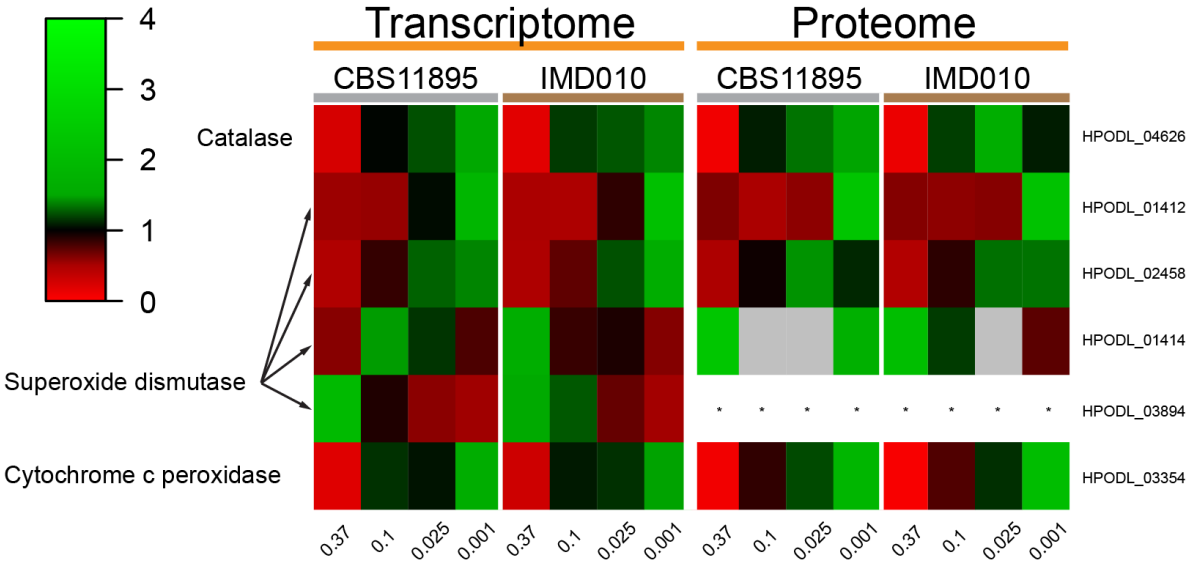


Figure S4.5: Mean-normalized transcript and protein abundances of genes related to detoxification of reactive oxygen species in *O. parapolymorpha* strains CBS11895 (wild type) and IMD010 (disrupted Complex I Nubm subunit). Samples were taken from duplicate independent aerobic glucose-grown batch (0.37 h^{-1}), chemostat (0.1 and 0.025 h^{-1}) and late-stage retentostat (0.001 h^{-1}) cultures. Transcript and protein abundances were mean-normalized separately for each gene and strain. Grey color: protein not detected based on criteria described in Methods section. Annotation of *O. parapolymorpha* catalase and superoxide dismutases was taken from (Ravin *et al.* 2013), and HPODL_03354 was identified as ortholog of *S. cerevisiae* cytochrome c peroxidase (*CCP1*) via the orthologous matrix database (see methods section). *HPODL_03894 was excluded from proteome analysis because abundance data did not pass quality requirements for analysis (see Methods section).

Chapter 5

Physiological relevance, localization and substrate specificity of the alternative (type II) mitochondrial NADH dehydrogenases of *Ogataea parapolymorpha*

Hannes Jürgens
Álvaro Mielgo-Gómez
Albert Godoy-Hernández
Jolanda ter Horst
Janine M. Nijenhuis
Duncan G. G. McMillan
Robert Mans

Essentially as submitted to Biorxiv (2021)
https://doi.org/10.1101/2021.04.28.441406

Abstract

Mitochondria from *Ogataea parapolymorpha* harbor a branched electron-transport chain containing a proton-pumping Complex I NADH dehydrogenase and three alternative (type II) NADH dehydrogenases (NDH2s). To investigate the physiological role, localization and substrate specificity of these enzymes, growth of various NADH dehydrogenase mutants was quantitatively characterized in shake-flask and chemostat cultures, followed by oxygen-uptake experiments with isolated mitochondria. Furthermore, NAD(P)H:quinone oxidoreduction of the three NDH2s were individually assessed. Our findings show that the *O. parapolymorpha* respiratory chain contains an internal NADH-accepting NDH2 (Ndh2-1/OpNdi1), at least one external NAD(P)H-accepting enzyme and likely additional mechanisms for respiration-linked oxidation of cytosolic NADH. Metabolic regulation appears to prevent competition between OpNdi1 and Complex I for mitochondrial NADH. With the exception of OpNdi1, the respiratory chain of *O. parapolymorpha* exhibits metabolic redundancy and tolerates deletion of multiple NADH-dehydrogenase genes without compromising fully respiratory metabolism.

Importance. To achieve high productivity and yields in microbial bioprocesses, efficient use of the energy substrate is essential. Organisms with branched respiratory chains can respire *via* the energy-efficient proton-pumping Complex I, or make use of alternative NADH dehydrogenases (NDH2s). The yeast *Ogataea parapolymorpha* contains three uncharacterized, putative NDH2s which were investigated in this work. We show that *O. parapolymorpha* contains at least one 'internal' NDH2, which provides an alternative to Complex I for mitochondrial NADH oxidation, albeit at a lower efficiency. The use of this NDH2 appeared to be limited to carbon excess conditions and the *O. parapolymorpha* respiratory chain tolerated multiple deletions without compromising respiratory metabolism, highlighting opportunities for metabolic (redox) engineering. By providing a more comprehensive understanding of the physiological role of NDH2s, including insights into their metabolic capacity, orientation and substrate specificity this study also extends our fundamental understanding of respiration in organisms with branched respiratory chains.

Introduction

Dissimilation of glucose to carbon dioxide results in the generation of reducing equivalents in the form of NADH, which are continuously (re)oxidized and intrinsically linked to the formation of ATP by the respiratory chain. In eukaryotes, glucose dissimilation and NADH generation occurs in the cytosol by glycolysis and in the mitochondrial matrix by the combined action of the pyruvate-dehydrogenase complex and the tricarboxylic acid cycle. NADH cannot cross the inner mitochondrial membrane (von Jagow and Klingenberg 1970), and as a result needs to be (re)oxidized in the cellular compartment in which it is generated.

Fungal respiratory chains are branched and contain multiple entry points for electrons from NADH. While the type I NADH:quinone oxidoreductase (NDH1 or Complex I) couples oxidation of mitochondrial NADH to the translocation of protons over the inner mitochondrial membrane, most fungi also possess so-called alternative NADH dehydrogenases (NDH2s) that catalyze NADH:quinone oxidoreduction without proton translocation (Joseph-Horne *et al.* 2001). These 'alternative NADH dehydrogenases' are monotopic proteins that attach to the inner mitochondrial membrane, but may differ in which side of the membrane they are located. Their catalytic sites either face the mitochondrial matrix ('internal') where their catalytic activity overlaps with Complex I, or the intermembrane space ('external'), allowing direct oxidation of cytosolic NADH (Antos-Krzeminska and Jarmuszkiewicz 2019, Feng *et al.* 2012, Melo *et al.* 2004). A prominent example for the phenotypical role of NDH2 can be found in the yeast *Saccharomyces cerevisiae*, which does not harbor a Complex I and instead relies solely on one internal and two external alternative NADH dehydrogenases as entry points for NADH-derived electrons into the respiratory chain (de Vries and Marres 1987, Luttik *et al.* 1998). Besides NADH, some external fungal NDH2s, such as the external alternative dehydrogenases from *Kluyveromyces lactis* (Tarrio *et al.* 2006, Tarrio *et al.* 2005) and *Neurospora crassa* (Carneiro *et al.* 2004, Carneiro *et al.* 2007, Melo *et al.* 2001), have also been reported to accept NADPH as substrate, either exclusively, or in addition to NADH. Based on the distribution of a key acidic residue (E272 in Ndi1 from *S. cerevisiae*), which has been proposed to prevent interaction with the phosphate group of NADPH (Iwata *et al.* 2012, Michalecka *et al.* 2004), it has been suggested that the majority of NDH2s oxidize NADH rather than NADPH (Marreiros *et al.* 2016b).

Ogataea parapolymorpha (formerly *Hansenula polymorpha*) is a methylotrophic, thermotolerant, Crabtree-negative yeast that is characterized by its rapid aerobic growth (Juergens *et al.* 2018a, Levine and Cooney 1973). It has a branched respiratory chain that contains Complex I and three putative alternative NAD(P)H dehydrogenases (referred to as Ndh2-1, Ndh2-2 and Ndh2-3 in this study) with unknown substrate

specificity and unknown orientation on the inner mitochondrial membrane (Juergens *et al.* 2020a). Interestingly in the presence of excess glucose and oxygen, deletion of Complex I from *O. parapolymorpha* does not result in a reduction of the maximum specific growth rate or biomass yield. This implies that Complex I is disposable under these conditions. In aerobic glucose-limited cultures, elimination of Complex I resulted in mutant strain which exhibited a 16% lower biomass yield while it maintained a fully respiratory metabolism (Juergens *et al.* 2020a). These phenotypes suggested that *O. parapolymorpha* harbors at least one internal alternative NADH dehydrogenase capable of compensating for the lack of Complex I, albeit at a lower efficiency of respiration-coupled ATP production. Such an internal enzyme also would likely be responsible for (re)oxidation of mitochondrial NADH under fast-growing glucose-excess conditions, especially since NDH2s appear to be suited to catalyze the high rates of NAD^+ regeneration required to sustain a high glycolytic flux in the absence of fermentation (Godoy-Hernandez *et al.* 2019).

The aim of this study was to investigate the physiological roles of the three putative alternative NAD(P)H dehydrogenases of *O. parapolymorpha* with a special focus on determining whether this yeast indeed possesses an internal alternative NADH dehydrogenase capable of functionally substituting Complex I. To this end, the aerobic growth characteristics of various *O. parapolymorpha* NAD(P)H-dehydrogenase mutant strains were investigated in glucose-grown batch- and chemostat cultures. To determine localization and substrate specificity of the alternative dehydrogenases, substrate-dependent oxygen-uptake rates of isolated mitochondria from wild-type and mutant strains were measured, and activity of the individual membrane-bound dehydrogenases with NADH and NADPH was assessed.

Results

Disruption of Ndh2-1 leads to a reduction of specific growth rate and a Crabtree-positive phenotype in *O. parapolymorpha*. To investigate the contribution to respiration of the three putative alternative NAD(P)H dehydrogenases, strains IMD003, IMD004 and IMD005 were constructed, which harbored a disrupted version of the structural gene for Ndh2-1, Ndh2-2 or Ndh2-3, respectively. The physiology of these mutant strains was then assessed in aerobic shake-flask batch cultures in the presence of excess glucose (2 g L^{-1}). The high specific rate of oxygen uptake by Crabtree-negative yeasts can lead to oxygen limitation in shake-flask cultures resulting in respiro-fermentative metabolism (Kiers *et al.* 1998). However, under the cultivation conditions in this study the wild-type *O. parapolymorpha* strain CBS11895 grows with a fully respiratory phenotype and at a comparable specific growth rate as previously described in fully aerobic bioreactors, indicating that oxygen limitation did not occur (**Table 5.1**) (Juergens *et al.* 2018a).

Table 5.1: Physiology of wild-type *O. parapolymorpha* CBS11895 and congenic mutant strains in aerobic shake-flask cultures grown at 30°C in synthetic medium with urea as nitrogen- and glucose (2 g L^{-1}) as carbon source. Data are presented as mean ± mean absolute deviation from at least two independent shake-flask cultures. Specific growth rates and ethanol yield were calculated from the exponential growth phase.

Strain (genotype)	Specific growth rate (h^{-1})	Physiology
CBS11895 (*wild type*)	0.36 ± 0.01	Respiratory
IMD003 (*Δndh2-1*)	0.30 ± 0.01	0.31 ± 0.02 mol_{EtOH} $mol_{Glucose}^{-1}$
IMD004 (*Δndh2-2*)	0.34 ± 0.01	Respiratory
IMD005 (*Δndh2-3*)	0.35 ± 0.01	Respiratory
IMX1978 (*Δndh2-2 Δndh2-3)*	0.32 ± 0.00	Respiratory
IMX2017 (*Δndh2-1 Δndh2-2 Δndh2-3)*	0.30 ± 0.00	0.29 ± 0.02 mol_{EtOH} $mol_{Glucose}^{-1}$
IMX2197 (*Δnubm Δndh2-2 Δndh2-3)*	0.31 ± 0.00	Respiratory

In shake-flask cultures, the specific growth rate of mutant strain IMD005 (*Δndh2-3*) did not differ significantly from that of the wild-type strain CBS11895, whereas strains IMD003 (*Δndh2-1*) and IMD004 (*Δndh2-2*) exhibited 17 and 6% lower specific growth rates, respectively (**Table 5.1**). In cultures of strains CBS11895, IMD004 and IMD005 no fermentation products were detected, indicating that a single disruption of *NDH2-2* or *NDH2-3* did not impede aerobic respiratory metabolism. In contrast, strain IMD003 exhibited a Crabtree-positive phenotype, producing 0.31 ± 0.02 mol of ethanol for each mol of glucose consumed. We then combined disruptions in *NDH2-2* and *NDH2-3*, resulting in strain IMX1978, which exhibited a slightly lower specific growth rate than

the strains with individually disrupted *NDH* genes, but still maintained a respiratory metabolism (**Table 5.1**). An additional disruption in strain IMX1978 of *NUBM*, which encodes an essential 51 kDA subunit of respiratory Complex I (Fecke *et al.* 1994), also did not result in detectable ethanol production in the resulting strain IMX2197 (*Δnubm Δndh2-2 Δndh2-3*). Collectively, these results demonstrate that Ndh2-1 presence alone suffices for supporting NAD(P)H turnover requirements in *O. parapolymorpha* for fully respiratory growth in glucose-grown batch cultures.

Strain IMX2017, in which all three *NDH2* genes were disrupted, exhibited a phenotype similar to that of strain IMD003 (*Δndh2-1*), displaying a 17% lower specific growth rate than observed for the wild-type strain CBS11895, and an ethanol yield similar to IMD003 (**Table 5.1**) Since yeast respiratory chains typically possess either none, or a single internal alternative NAD(P)H dehydrogenase (Antos-Krzeminska and Jarmuszkiewicz 2019), and respiratory Complex I is not physiologically relevant under these conditions in *O. parapolymorpha* (Juergens *et al.* 2020a), these data support the hypothesis that Ndh2-1 has an internal orientation (i.e. facing the mitochondrial matrix).

Oxygen consumption studies with mitochondria from wild-type *O. parapolymorpha* and deletion mutants confirm internal orientation of Ndh2-1. Measurement of the oxygen-uptake rates of isolated mitochondria using a compartmentalized substrate approach has been previously used to determine the orientation of yeast mitochondrial NAD(P)H dehydrogenases (Gonzalez-Barroso *et al.* 2006, Luttik *et al.* 1998, Tarrio *et al.* 2005). To obtain *O. parapolymorpha* mitochondria, we adapted a protocol for isolation of mitochondria from glucose-limited *S. cerevisiae* cultures (Luttik *et al.* 1998). Initial isolations from wild-type cells grown in glucose-limited chemostat cultures resulted in well-coupled *O. parapolymorpha* mitochondria with a 'respiratory control ratio' (RCR) of 3.6 when assayed with endogenous NADH (generated in the mitochondrial matrix by addition of pyruvate and malate). However, the same preparations exhibited rapid, uncoupled oxygen uptake when exposed to methanol or ethanol. Methanol- and ethanol-dependent oxygen-uptake rates were approximately 4-fold and 2.5-fold faster, respectively, than ADP-stimulated respiration of endogenous NADH. (M)ethanol-dependent oxygen uptake indicated a potential contamination of the mitochondrial preparations of these glucose-derepressed cultures with peroxisomes containing methanol oxidase (MOX) (Egli *et al.* 1980, Yurimoto *et al.* 2011). A similar contamination was previously observed in mitochondrial preparations from the methylotrophic yeast *Pichia pastoris* (Gonzalez-Barroso *et al.* 2006). In an attempt to obtain mitochondrial preparations devoid of MOX activity, mitochondria were also isolated from cells grown under MOX-repressing conditions in glucose-grown batch- or nitrogen-limited chemostat cultures (Egli and Quayle 1986, Ravin *et al.* 2013).

Mitochondrial isolations from cells grown under these conditions did not consume oxygen in the presence of methanol, but they exhibited RCRs close to 1 when assayed with endogenous NADH, indicating uncoupled preparations (data not shown). In light of these findings, mitochondria isolated from glucose-limited chemostat cultures were used for respiration studies, and interference of the oxygen-uptake measurements by MOX activity was minimized by using reaction mixtures and substrates devoid of alcoholic solvents (see Methods section).

To test the hypothesis that the catalytic site of Ndh2-1 is oriented towards the mitochondrial matrix, oxygen uptake was measured in the presence of endogenous and exogenous NADH using mitochondria isolated from wild-type *O. parapolymorpha* (CBS11895), from strains possessing only Complex I (IMX2017; *NUBM Δndh2-1 Δndh2-2 Δndh2-3*) or only Ndh2-1 (IMX2197; *Δnubm NDH2-1 Δndh2-2 Δndh2-3*) as known respiration-linked NAD(P)H dehydrogenases (**Table 5.2**). These strains were of specific interest because in aerobic glucose-limited chemostats at a dilution rate of 0.1 h^{-1}, all three strains exhibited a fully respiratory metabolism (see below, **Table 5.3**), indicating a functional respiratory chain responsible for (re)oxidation of both cytosolic and mitochondrial NADH.

Mitochondria isolated from strain CBS11895 readily consumed oxygen in the presence of endogenous and exogenous NADH (**Table 5.2**). RCRs for endogenous NADH (3.6 ± 0.4) and exogenous NADH (2.5 ± 0.1) indicated that the observed respiration-linked NADH oxidation is due to the activity of separate internal and external NADH dehydrogenases and not the result of physically compromised mitochondria. Furthermore, the near-complete inhibition of oxygen utilization by the cytochrome c oxidase inhibitor cyanide (~92 and ~94% effective in the presence of endogenous and exogenous NADH, respectively) strongly suggests that oxygen consumption was dependent due to membrane-bound oxidative phosphorylation. In contrast to CBS11895, mitochondria from strains IMX2017 and IMX2197 both exhibited 88% lower oxygen consumption rates in the presence of exogenous NADH, suggesting that mitochondria from these strains do not possess external NADH dehydrogenase activity.

When assayed with endogenous NADH, mitochondria from strains IMX2017 ('Complex I only') and IMX2197 ('Ndh2-1 only') exhibited similar coupled oxygen-uptake rates to mitochondria isolated from the wild-type strain CBS11895 (**Table 5.2**). Presence of the specific Complex I inhibitor rotenone strongly decreased oxygen-uptake rates with endogenous NADH of mitochondria from the strains CBS11895 and IMX2017 (by 80 and 71%, respectively), while for strain IMX2197 no significant rotenone inhibition was observed. These observations demonstrate that Ndh2-1 is indeed internally oriented and able to completely take over the role of Complex I in NADH oxidation in *Δnubm* mutants under aerobic, glucose-limited conditions.

When mitochondria isolated from the wild-type strain *O. parapolymorpha* CBS11895 were assayed with exogenous NADPH, oxygen consumption was detected at a rate of 0.09 ± 0.01 µmol O_2 (mg protein)$^{-1}$ min^{-1}, substantially lower than the rate obtained with exogenous NADH (0.33 ± 0.05 µmol O_2 (mg protein)$^{-1}$ min^{-1}, **Table 5.2**). NADPH-dependent oxygen consumption was not significantly coupled (RCR of 1.1). However, it was largely inhibited by cyanide (~85%), indicating that NADPH oxidation did not occur *via* a soluble enzyme but by an external NADPH-accepting dehydrogenase that transferred electrons from NADPH into the mitochondrial respiratory chain. In strains IMX2017 and IMX2197, exogenous NADPH oxidation activity was reduced by 3- and 9-fold, respectively, indicating the NADPH oxidation observed in CBS11895 is caused by Ndh2-2 and/or Ndh2-3.

Table 5.2: Substrate-dependent rates of oxygen uptake by mitochondria from wild-type *O. parapolymorpha* (CBS11895) and mutants possessing only Complex I (IMX2017) or Ndh2-1 (IMX2197) as known respiration-linked NAD(P)H dehydrogenases. Mitochondria were isolated from cells grown in aerobic, glucose-limited chemostat cultures at a dilution rate (D) of 0.1 h^{-1} and assayed at 30°C, pH 7.0. Oxygen uptake in µmol O_2 (mg protein)$^{-1}$ min^{-1} was determined in the presence of 0.25 mM ADP. Respiratory control ratio (RCR) values represent the ratio of oxygen uptake rates in the presence and absence (prior to addition) of ADP. For tests with rotenone (50 µM), mitochondria were pre-incubated with this inhibitor at assay conditions prior to substrate addition. Oxygen-uptake rates and RCRs are presented as mean ± standard deviation of measurements with mitochondria from at least two independent chemostat cultures for each strain. Tests with KCN (1 mM) were performed with mitochondria from a single, representative isolation of strain CBS11895. N.D., not determined.

Substrate	CBS11895 (*NUBM NDH2-1 NDH2-2 NDH2-3*)			IMX2017 (*NUBM Δndh2-1 Δndh2-2 Δndh2-3*)		IMX2197 (*Δnubm NDH2-1 Δndh2-2 Δndh2-3*)	
	O_2 uptake rate	**RCR**	**KCN inhibition**	**O_2 uptake rate**	**RCR**	**O_2 uptake rate**	**RCR**
Pyruvate + malate	0.15 ± 0.02	3.6 ± 0.4	92%	0.14 ± 0.00	3.2 ± 0.4	0.18 ± 0.00	2.7 ± 0.6
Pyruvate + malate (rotenone)	0.03 ± 0.00	1.6 ± 0.0	N.D.	0.04 ± 0.00	1.9 ± 0.1	0.16 ± 0.00	3.1 ± 0.5
NADH	0.33 ± 0.05	2.5 ± 0.1	94%	0.04 ± 0.00	1.4 ± 0.2	0.04 ± 0.01	1.1 ± 0.0
NADPH	0.09 ± 0.01	1.1 ± 0.0	85%	0.03 ± 0.00	2.9 ± 0.3	0.01 ± 0.00	1.4 ± 0.2

***O. parapolymorpha* mutants with linearized respiratory chains exhibit respiratory physiology in glucose-limited chemostat cultures.** In aerobic glucose-limited chemostat cultures grown at a dilution rate of 0.1 h^{-1}, both strain IMX2017 (*NUBM Δndh2-1 Δndh2-2 Δndh2-3*) and strain IMX2197 (*Δnubm NDH2-1 Δndh2-2 Δndh2-3*) exhibited essentially the same respiratory physiology as the wild-type strain CBS11895. Despite the deletion of genes encoding respiration-linked NAD(P)H dehydrogenases, both mutant strains grew without detectable formation of fermentation products. Moreover, an oxidative respiratory quotient close to 1 was observed and the carbon contained in the glucose feed could be completely recovered

as biomass and carbon dioxide (**Table 5.3**). IMX2017 exhibited a biomass yield of 0.52 g biomass (g glucose)$^{-1}$, which is not significantly different from that found with wild type CBS11895 (0.51 g biomass (g glucose)$^{-1}$). In contrast, the biomass yield of strain IMX2197 was reduced by ~15% compared to that of the two other strains. This reduced biomass yield is consistent with a less efficient oxidative respiratory chain, using internal, non-proton pumping Ndh2-1 instead of the proton-pumping Complex I. Accordingly, IMX2197 exhibited an approximately ~30% lower biomass yield on oxygen, with correspondingly higher biomass-specific rates of oxygen consumption and carbon-dioxide production.

Table 5.3: Physiology of *Ogataea parapolymorpha* strains CBS11895, IMX2017 and IMX2197 in aerobic, glucose-limited chemostat cultures grown at a dilution of 0.1 h^{-1}, 30°C and pH 5. Data are presented as mean ± mean absolute deviation from two independent replicates. Carbon recoveries were calculated based on a biomass carbon content of 48% (w/w). Symbols: $Y_{X/S}$ and $Y_{X/O2}$ = yield of biomass dry weight on glucose and oxygen, respectively; RQ = respiratory quotient; $q_{Glucose}$, q_{CO2} and q_{O2} represent biomass-specific uptake/production rates of glucose, CO_2 and O_2, respectively; C_X represents biomass dry weight concentration. *Data for CBS11895 were reproduced from Juergens *et al.* (2018a).

Strain	CBS11895* wild type	IMX2017 only Complex I	IMX2197 only Ndh2-1
Actual dilution rate (h^{-1})	0.10 ± 0.00	0.10 ± 0.00	0.10 ± 0.00
Reservoir glucose (g L^{-1})	7.49 ± 0.00	7.36 ± 0.01	9.04 ± 0.01
$Y_{X/S}$ (g biomass [g glucose]$^{-1}$)	0.51 ± 0.00	0.52 ± 0.00	0.44 ± 0.01
$Y_{X/O2}$ (g biomass [g O_2]$^{-1}$)	1.35 ± 0.05	1.36 ± 0.00	0.94 ± 0.01
RQ (-)	1.05 ± 0.01	1.05 ± 0.00	1.04 ± 0.00
$q_{Glucose}$ (mmol [g biomass]$^{-1}$ h^{-1})	-1.08 ± 0.03	-1.07 ± 0.01	-1.28 ± 0.00
q_{CO2} (mmol [g biomass]$^{-1}$ h^{-1})	2.44 ± 0.07	2.40 ± 0.01	3.50 ± 0.00
q_{O2} (mmol [g biomass]$^{-1}$ h^{-1})	-2.32 ± 0.08	-2.30 ± 0.01	-3.38 ± 0.00
C_X (g biomass L^{-1})	3.84 ± 0.08	3.83 ± 0.04	4.01 ± 0.05
Carbon recovery (%)	99.3 ± 1.2	100.1 ± 0.8	98.8 ± 0.8

The *O. parapolymorpha* NDH2s oxidize NADH but not NADPH when expressed in *E. coli* membranes. The oxygen-consumption experiments indicated that mitochondria from *O. parapolymorpha* poorly couple oxidation of exogenous NADPH to oxygen consumption *via* the aerobic respiratory chain. In principle, any external NDH2 could be responsible for this activity. Upon closer examination of the putative amino acid sequence of *O. parapolymorpha* Ndh2-3, an uncharged residue (Q365) instead of a negatively charged (E272 in *S. cerevisiae* Ndi1) is present within the substrate-binding domain (**Figure S5.1**). This residue has been suggested to be involved in determining NADH/NADPH specificity due to interaction with the phosphate group of NADPH (Iwata *et al.* 2012, Michalecka *et al.* 2004).

To assess the ability of the *O. parapolymorpha* NDH2s to catalyze the oxidation of NADH and/or NADPH, Ndh2-1, Ndh2-2 and Ndh2-3 were individually overexpressed in *Escherichia coli*, following a strategy previously applied to an NDH2 (Ndi1) from *S. cerevisiae* (Kitajima-Ihara and Yagi 1998). Since respiration-linked NADPH-dehydrogenase activity has not been reported for *E. coli*, this host was regarded as especially suitable to assess NADPH-dehydrogenase activity of heterologously expressed enzymes.

In spectrophotometric assays at pH 7.4, expression of each of the three *O. parapolymorpha* NDH2s led to 2.4 to 3.2-fold higher NADH-oxidation rates than observed with membranes isolated from the parental *E. coli* strain (**Figure 5.1**). This activity indicated successful overexpression and localization to the *E. coli* membrane, and confirmed that the three *O. parapolymorpha* NDH2s are indeed NADH:quinone oxidoreductases. When NADPH was added as substrate to the same membrane preparations, no detectable NADH-oxidation was measurable, indicating that none of the three *O. parapolymorpha* NDH2s can effectively utilize NADPH under the conditions tested (**Table S5.1**). Since NAD(P)H oxidation by NDH2s can depend on pH (Melo *et al.* 2001, Nantapong *et al.* 2005), NADH and NADPH oxidation measurements were repeated at pH 5.5 and 8.0. These different pH values did not influence NADH oxidation rates relative to rates measured at pH 7.4 (Student's *t* test, $p > 0.05$), and did not stimulate NADPH utilization by either endogenous *E. coli* respiratory enzymes or *O. parapolymorpha* NDH2s (**Table S5.1**). Finally, oxidative NAD(P)H catalysis by Ndh2-3 overexpressed in *E. coli* was also tested in the presence of calcium, as Ndh2-3 contains a putative EF-hand suggesting a calcium binding domain which could potentially regulate catalytic behavior (**Figure S5.2**) (Melo *et al.* 2001). However, the presence of 5 mM Ca^{2+} did not significantly affect rates of NADH oxidation by Ndh2-3 overexpression membranes at pH 5.5 and 7.4 (24% reduction at pH 8.0) and did not enable NADPH utilization at pH 5.5, 7.4 or 8.0 (**Table S5.1**).

Taken together, these data conclusively show that when assessed in *E. coli* membranes, the three NDH2s from *O. parapolymorpha* can efficiently oxidize NADH at a range of pH levels (pH 5.5-8.0), but do not catalyze the oxidation of NADPH.

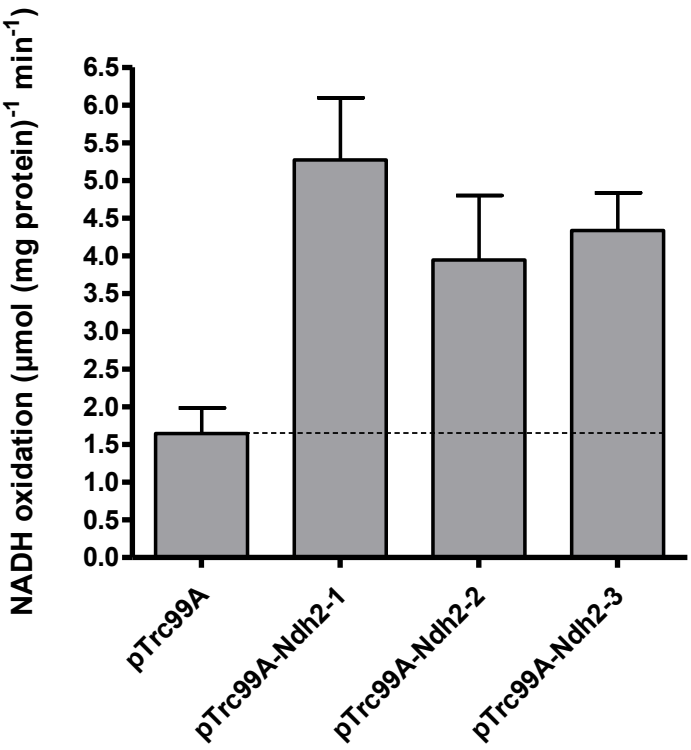


Figure 5.1: NADH oxidation by *E. coli* membranes isolated from strains overexpressing individual *O. parapolymorpha* NDH2s. Control measurements were done with membranes isolated from a strain carrying an empty overexpression plasmid (pTrc99A). Assays were performed with a membrane protein concentration of 10 µg mL^{-1} at 37°C, with 200 µM NADH and 100 µM ubiquinone-1. Data is presented as mean ± standard deviation of triplicate measurements.

Discussion

NAD(P)H dehydrogenases in the respiratory chain of *O. parapolymorpha*. In this study, the *O. parapolymorpha* gene HPODL_02792 (*NDH2-1*) was demonstrated to encode a mitochondrial 'internal' alternative NADH dehydrogenase. To achieve consistent nomenclature with other fungal species, we suggest naming this gene *OpNDI1* (NADH dehydrogenase, internal). The closely related yeast *O. polymorpha* likely also possesses an internal alternative NADH dehydrogenase, as its genome (Riley *et al.* 2016) encodes a protein (OGAPODRAFT_15258) that exhibits 98% sequence identity with OpNdi1.

Mitochondria isolated from *O. parapolymorpha* also oxidized externally supplied NADH, indicating that either one or both of the remaining alternative dehydrogenases, Ndh2-2 and Ndh2-3, are facing the intermembrane space, rather than the mitochondrial matrix. Coexistence of more than one internal alternative dehydrogenase has been described (Rasmusson *et al.* 2004), but since the fungi and yeasts so far characterized either possess one or no such enzymes, both Ndh2-2 and Ndh2-3 appear to be externally oriented. The internal orientation of Ndh2-1/OpNdh1 has now been established in this study, oxygen uptake experiments with mitochondria isolated from strains IMD004 (*Δndh2-2*) and IMD005 (*Δndh2-3*) could be performed to confirm the localization of the other two enzymes.

Wild-type *O. parapolymorpha* mitochondria also oxidized NADPH *via* the respiratory chain as demonstrated by near-complete inhibition of this activity by cyanide. While growth on glucose the production of NADPH can be balanced with biosynthetic requirements, assimilation of other carbon sources such as gluconate result in a surplus of NADPH in yeast (Bruinenberg *et al.* 1983b), making the ability to respire NADPH a beneficial trait. The oxygen-uptake activity with NADPH was essentially uncoupled (RCR of 1.1), which was unexpected, as addition of ADP should relieve the backpressure of the proton gradient, especially since the same mitochondrial preparations exhibited well-coupled oxygen uptake in the presence of NADH. Mitochondria isolated from glucose-limited chemostat cultures of *K. lactis* (D = 0.1 h^{-1}) similarly exhibited uncoupled oxygen uptake in the presence of NADPH (Overkamp *et al.* 2002), while mitochondria from lactate- and glucose-grown batch cultures of the same yeast exhibited RCRs of 2.3 and 1.3 for NADPH, respectively (Tarrio *et al.* 2006). With mitochondria isolated from glucose-limited chemostats of *Candida utilis*, the RCRs of NADPH oxidation were found to vary as a function of the dilution rate, with cultures grown at D = 0.05-0.1 h^{-1} exhibiting an RCR of ~1.9 while cultures grown at D = 0.2-0.4 h^{-1} exhibited a lower RCR of ~1.2 (Bruinenberg *et al.* 1985b). In agreement with our study, these studies reported inhibition of NADPH oxidation by cyanide and/or antimycin A, indicating that the degree of oxidative coupling of NADPH respiration by

yeast mitochondria is dependent on the utilized substrate and/or the growth condition employed, which might also be the case in *O. parapolymorpha*.

The heterologous expression of *O. parapolymorpha* NDH2s in *E. coli* membranes showed that NADPH oxidation of the *O. parapolymorpha* mitochondria is not likely to have occurred *via* these NDH2(s), since the three enzymes exclusively utilized NADH when assayed within *E. coli* membranes. Based on their protein sequences, we originally speculated Ndh2-3 to be the most likely candidate to accept NADPH, as it contains the uncharged residue proposed to permit NADPH utilization (Michalecka *et al.* 2004). There are indeed NADPH-utilizing NDH2s with this uncharged residue such as Nde1 from *N. crassa* (**Figure S5.1**) or plant enzyme St-NBD1 (Michalecka *et al.* 2004), and mutation of this exact residue has been exploited to alter substrate specificity from NADH to co-utilization of NADH and NADPH in a bacterial NDH2 (Desplats *et al.* 2007). However, Nde2 from *N. crassa* and Nde1 from *K. lactis* do not contain the uncharged residue but still accept both substrates (**Figure S5.1**), indicating that a charged amino acid in this position does not strictly prevent NADPH utilization and that NDH2 substrate specificity cannot be accurately predicted by a single residue. In contrast to Nde1 from *N. crassa*, where NADPH oxidation is highly affected by calcium (Melo *et al.* 2001), the putative calcium binding domains of Ndh2-3 from *O. parapolymorpha* and Nde2 from *K. lactis* are poorly conserved, and apparently lost their regulatory function, as presence of calcium did not appear to have an effect on activity of either enzyme (Tarrio *et al.* 2006).

NDH2-independent mechanisms for respiration of cytosolic NADH. Mitochondria from strains IMX2017 (*Δndh2-1 Δndh2-2 Δndh2-3*) and IMX2197 (*Δnubm Δndh2-2 Δndh2-3*) did not oxidize external NADH, but both strains exhibited a fully respiratory phenotype in glucose-limited chemostat cultures at D = 0.1 h^{-1}. Since both strains only contain a NADH dehydrogenase that can accept electrons from NADH in the mitochondrial matrix (Complex I and Ndh2-1/OpNdi1, respectively), mechanisms other than the external alternative NADH dehydrogenases are capable of respiring cytosolic NADH in these strains must be present. One candidate would be the 'Gut2/Gpd shuttle', consisting of mitochondrial glycerol-3-phosphate dehydrogenase (Gut2), which indirectly respires cytosolic NADH in combination with cytosolic NAD-dependent glycerol 3-phosphate dehydrogenase (Gpd) (Bakker *et al.* 2001). When grown on glycerol, *O. parapolymorpha* exhibits highly increased glycerol-kinase activity, indicating that glycerol assimilation occurs *via* the phosphorylative pathway. This indicates that Gut2 is functional in this yeast (Tani and Yamada 1987). Furthermore, Gut2 activity was demonstrated in the closely related yeast *O. polymorpha* (de Koning *et al.* 1987). In *S. cerevisiae*, the metabolic function of Gut2 overlaps with the external NDH2s (Larsson *et al.* 1998, Overkamp *et al.* 2000), and yeast mitochondria

isolated from glucose-grown cultures of various yeast species have been demonstrated to respire glycerol-3-phosphate with similar specific oxygen uptake rates as NADH (Gonzalez-Barroso *et al.* 2006, Luttik *et al.* 1998, Overkamp *et al.* 2002).

While *S. cerevisiae* lacking Gut2 in addition to its two external NADH dehydrogenases produces large amounts of glycerol under aerobic, glucose-limited conditions (Overkamp *et al.* 2000), an *O. parapolymorpha* mutant lacking Gut2 in addition to the three NDH2s (*Δndh2-1 Δndh2-2 Δndh2-3 Δgut2*) exhibited fully respiratory physiology under identical conditions (**Table S5.2**). The absence of byproduct formation indicates that at least one additional, unknown mechanism for coupling the oxidation of cytosolic NADH to mitochondrial respiration is present in *O. parapolymorpha*. Such a mechanism could, for example, involve a shuttle for import of NADH equivalents into the mitochondrial matrix, where oxidation by Complex I and/or OpNdi1 can occur. For *S. cerevisiae*, several suitable mechanisms have been demonstrated or proposed, such as the malate-oxaloacetate and malate-aspartate shuttles, mitochondrial oxidation of ethanol produced in the cytosol, or, assuming a high cytosolic NADH/NAD^+ ratio, an inward-directed ethanol-acetaldehyde shuttle (Bakker *et al.* 2001, Palmieri *et al.* 2006). Based on sequence homology with *S. cerevisiae*, key proteins for these mechanisms are present in *O. parapolymorpha* (**Table S5.3**), however their operation and physiological relevance has not been confirmed by functional analysis studies

Physiological relevance of branched respiratory chains. Respiratory chains of plants, fungi and some protists are branched (Antos-Krzeminska and Jarmuszkiewicz 2019, Rasmusson *et al.* 2008), and, as also demonstrated in this study for *O. parapolymorpha*, can exhibit a metabolic redundancy for respiration-linked NADH oxidation, minimizing the physiological effect of loss of individual or multiple NADH dehydrogenases (Carneiro *et al.* 2004, Fromm *et al.* 2016, Overkamp *et al.* 2000). In general, this redundancy and the exact physiological role of the alternative NADH dehydrogenases are still poorly understood. In *O. parapolymorpha*, OpNdi1 appears to have a unique metabolic function as it is the only NADH dehydrogenase strictly required to sustain fast respiratory growth under glucose excess conditions, demonstrating that limiting respiratory capacity by disruption of a single NADH dehydrogenase can elicit the Crabtree effect in a Crabtree-negative yeast. Similarly, the overexpression of a single 'Gal4-like' transcription factor has been reported to convert Crabtree-negative *P. pastoris* into a Crabtree-positive yeast (Ata *et al.* 2018). However, since formation of ethanol and CO_2 from glucose by alcoholic fermentation is redox-neutral and reduced by-products such as glycerol were not detected in fermenting cultures of IMD003 (*Δndh2-1*) and IMX2017 (*Δndh2-1 Δndh2-2 Δndh2-3*), (re)oxidation of mitochondrial NADH from oxidative sugar metabolism must still occur via respiration. While

Complex I has been demonstrated to be physiologically irrelevant under these conditions in wild-type *O. parapolymorpha* CBS11895 (Juergens *et al.* 2020a), based on the available data its contribution to oxidation of mitochondrial NADH cannot be excluded in these strains. Alternatively, if export of NADH equivalents from the mitochondrial matrix is possible in this yeast, NADH could also be oxidized in the cytosol, for example by external NADH dehydrogenase(s) in IMD003 or the Gut2/Gpd shuttle in IMX2017.

A previous study with submitochondrial particles harvested from stationary-phase cultures of *O. polymorpha* found that most NADH oxidation occurred via NDH2 and only ~10% via Complex I (Bridges *et al.* 2009). Similarly, in our experiments with mitochondria isolated from wild-type *O. parapolymorpha* grown in glucose-limited chemostat cultures, only ~30% of the total specific NADH oxidation activity could be attributed to Complex I. However, due to the saturating substrate concentrations used in these types of experiments, they allow only for a limited interpretation of the actual physiological relevance of the respective enzymes. *In vivo*, competition of Complex I and OpNdi1 for NADH likely occurs if both systems are expressed at the same time, and indeed some fungal species appear to co-utilize Complex I and internal NDH2 (Fecke *et al.* 1994, Prömper *et al.* 1993, Voulgaris *et al.* 2012). With mitochondria isolated from wild-type *O. parapolymorpha* strain CBS11895, the presence of 50 µM rotenone did not fully inhibit internal NADH oxidation. However, as strain IMX2017, which lacked all NADH dehydrogenases besides Complex I, exhibited a similar partial inhibition of oxygen uptake by rotenone, the uninhibited activity in CBS11895 was likely not caused by OpNdi1 but instead by incomplete inhibition of *O. parapolymorpha* Complex I. Comparative studies with submitochondrial particles have demonstrated that rotenone inhibits Complex I from yeasts less strongly than the *Bos taurus* enzyme, requiring 50 µM rotenone to achieve 96% inhibition of NADH oxidation activity by the *P. pastoris* Complex I (Bridges *et al.* 2009). It is conceivable that Complex I from O*. parapolymorpha* is even more resistant to rotenone, explaining the observed partial inhibition in strains CBS11895 and IMX2017. In addition, OpNdi1 was not detected in the proteome of aerobic, glucose-limited cultures of CBS11895 (Juergens *et al.* 2020a), and the observed identical biomass yields of strains CBS11895 and IMX2017 are consistent with a situation in which essentially all mitochondrial NADH is (re)oxidized by Complex I in wild-type *O. parapolymorpha* under these conditions. These observations indicate that oxidation of mitochondrial NADH is strictly separated between Complex I and OpNdi1 in *O. parapolymorpha* under conditions of glucose limitation and glucose excess, respectively. Nevertheless, OpNdi1 is able to fully support respiratory growth in the absence of a functional Complex I, as previously hypothesized (Juergens *et al.* 2020a).

Concluding remarks. In this study we show that the Crabtree-negative yeast *O. parapolymorpha* contains both NDH1- and NDH2-type NADH dehydrogenases for respiration of NADH in the mitochondrial matrix, but limits their utilization to conditions of carbon limitation and carbon excess, respectively. Furthermore, we find that the respiratory chain of *O. parapolymorpha* can tolerate multiple deletions without compromising respiratory metabolism, offering insight into its flexible nature and opportunities for metabolic (redox) engineering in this industrially-relevant yeast. Finally, the phenotype elicited by disruption of OpNdi1 demonstrates that limiting respiratory capacity by a single mutation in an NADH dehydrogenase can result in overflow metabolism and convert *O. parapolymorpha* into a yeast with a Crabtree-positive phenotype.

Materials and Methods

Yeast strains and maintenance. The *O. parapolymorpha* strains used in this study are derived from the wild-type strain CBS11895 (DL-1; ATCC26012) and are described in **Table 5.4**. For construction and maintenance, yeast strains were grown in an Innova shaker incubator (New Brunswick Scientific, Edison, NJ, USA) set to 30°C and 200 rpm, in 500 mL shake flasks containing 100 mL heat-sterilized (120°C for 20 min) YPD medium (10 g L^{-1} Bacto yeast extract, 20 g L^{-1} Bacto peptone, 20 g L^{-1} glucose, demineralized water). Solid medium was prepared by addition of 2% (w/v) agar. Frozen stock cultures were prepared from exponentially growing shake-flask cultures by addition of glycerol to a final concentration of 30% (v/v), and aseptically stored in 1 mL aliquots at -80°C.

Table 5.4: *O. parapolymorpha* (*Hansenula polymorpha*) strains used in this study. [A]CBS11895 was obtained from the CBS-KNAW fungal collection (Westerdijk Fungal Biodiversity Institute, Utrecht, The Netherlands).

Strain	Genotype	Origin
CBS11895	Wild type *O. parapolymorpha* (DL-1)	CBS-KNAW[A]
IMD003	*ndh2-1$^{\Delta 266-269}$*	This study
IMD004	*ndh2-2^{A289AT}*	This study
IMD005	*ndh2-3$^{\Delta 505-515}$*	This study
IMX1945	*Δndh2-2::AgTEF1p-hph-AgTEF1t*	This study
IMX1978	*Δndh2-2::AgTEF1p-hph-AgTEF1t Δndh2-3::AaTEF1p-NatR-ScPHO5t*	This study
IMX2197	*Δndh2-2::AgTEF1p-hph-AgTEF1t Δndh2-3::AaTEF1p-NatR-ScPHO5t Δnubm::AgTEF1p-pat-AgTEF1t*	This study
IMX2017	*Δndh2-2::AgTEF1p-hph-AgTEF1t Δndh2-3::AaTEF1p-NatR-ScPHO5t Δndh2-1::ScTEF1p-kanR-ScTDH1t*	This study
IMX2167	*Δndh2-2::AgTEF1p-hph-AgTEF1t Δndh2-3::AaTEF1p-NatR-ScPHO5t Δndh2-1::ScTEF1p-kanR-ScTDH1t Δgut2::AgTEF1p-pat-AgTEF1t*	This study

Plasmid construction. All plasmids used in this study are described in **Table 5.5**. Plasmids pUD546, pUD547 and pUD548 were *de novo* synthesized by GeneArt (Thermo Fisher Scientific, Waltham, MA, USA) and contained synthetic guide RNA (gRNA) constructs with spacer sequences (5'-3') 'CCTGATGTAAATATACGCTG', 'AAGAAGAACATTGTTATTCT', and 'GTTATTCTGGGTTCCGGCTG', respectively. Cas9/gRNA co-expression plasmids pUDP019, pUDP020 and pUDP021 were constructed using pUD546, pUD547 and pUD548, respectively, by integration into pUDP002 via BsaI-mediated 'Golden Gate' assembly (Engler *et al.* 2008) as described previously (Juergens *et al.* 2018b), and verified by digestion with PdmI.

Table 5.5: Plasmids used in this study. Restriction enzyme sites are indicated in superscript, and genes targeted by gRNAs or homologous recombination are indicated in subscript. Aa: *Arxula adeninivorans*; Sp: *Streptococcus pyogenes*; Ag: *Ashbya gossypii*; Sc: *Saccharomyces cerevisiae*; HH: hammerhead ribozyme; HDV: hepatitis delta virus ribozyme; MCS: multiple cloning site; HRS: homologous recombination sequence.

Name	Relevant characteristics	Origin
pUD546	^{BsaI}HH-gRNA$_{NDH2\text{-}1}$-HDV^{BsaI}	GeneArt
pUD547	^{BsaI}HH-gRNA$_{NDH2\text{-}2}$-HDV^{BsaI}	GeneArt
pUD548	^{BsaI}HH-gRNA$_{NDH2\text{-}3}$-HDV^{BsaI}	GeneArt
pUDP002	*panARS (OPT) AgTEF1p-hph-AgTEF1t ScTDH3p*$^{BsaI\ BsaI}$*ScCYC1t AaTEF1p-Spcas9 -ScPHO5t*	(Juergens *et al.* 2018b)
pUDP019	*panARS (OPT) AgTEF1p-hph-AgTEF1t ScTDH3p*-HH-gRNA$_{NDH2\text{-}1}$-HDV-*ScCYC1t AaTEF1p-Spcas9 -ScPHO5t*	This study
pUDP020	*panARS (OPT) AgTEF1p-hph-AgTEF1t ScTDH3p*-HH-gRNA$_{NDH2\text{-}2}$-HDV-*ScCYC1t AaTEF1p-Spcas9 -ScPHO5t*	This study
pUDP021	*panARS (OPT) AgTEF1p-hph-AgTEF1t ScTDH3p*-HH-gRNA$_{NDH2\text{-}3}$-HDV-*ScCYC1t AaTEF1p-Spcas9 -ScPHO5t*	This study
pUD602	Template for plasmid backbone (*ori ampR*)	(Juergens *et al.* 2018b)
pYTK013	Template for *ScTEF1p* promoter	(Lee *et al.* 2015)
pYTK056	Template for *ScTDH1t* terminator	(Lee *et al.* 2015)
pYTK077	Template for *kanR* ORF	(Lee *et al.* 2015)
pYTK078	Template for *NatR* ORF	(Lee *et al.* 2015)
pAG31	Template for *AgTEF1p-pat-AgTEF1t* cassette	(Goldstein and McCusker 1999)
pUD740	Template for *ScTEF1p-kanR-ScTDH1t* cassette	This study
pUD801	HRS$_{NDH2\text{-}1}$-*ScTEF1p-kanR-ScTDH1t*-HRS$_{NDH2\text{-}1}$	This study
pUD802	HRS$_{NDH2\text{-}2}$-*AgTEF1p-hph-AgTEF1t* -HRS$_{NDH2\text{-}2}$	This study
pUD803	HRS$_{NDH2\text{-}3}$-*AaTEF1p-NatR-ScPHO5t*-HRS$_{NDH2\text{-}3}$	This study
pUD1035	HRS$_{GUT2}$-*AgTEF1p-pat-AgTEF1t*-HRS$_{GUT2}$	This study
pUD1036	HRS$_{NUBM}$-*AgTEF1p-pat-AgTEF1t*-HRS$_{NUBM}$	This study
pTrc99A	*ori ampR trp/lac*-MCS-*rrnB*	(Amann *et al.* 1988)
pTrc99A-NDH2-1	*ori ampR trp/lac-NDH2-1-rrnB*	This study
pTrc99A-NDH2-2	*ori ampR trp/lac-NDH2-2-rrnB*	This study
pTrc99A-NDH2-3	*ori ampR trp/lac-NDH2-3-rrnB*	This study

Construction of pUD740 was done via 'Gibson assembly' (Gibson *et al.* 2009) from the following PCR-amplified fragments: the *ScTEF1* promoter from pYTK013 (primers 12099 + 12100), the *kanR* G418 resistance marker ORF from pYTK077 (primers 12097 + 12098), the *ScTDH1* terminator from pYTK056 (primers 12095 + 12096), the *AaTEF1p-Spcas9-ScPHO5t* expression cassette from pUDP002 (primers 10426 + 10427), upstream (primers 12093 + 12094) and downstream (primers 12101 + 12102) homologous recombination sequences of the *OpKU70* locus from CBS11895 genomic DNA and the plasmid backbone from pUD602 (primers 12103 + 12104).

Plasmids pUD801, pUD802, pUD803, pUD1035 and pUD1036 carrying subcloned homology-flanked marker cassettes for split-marker deletion were constructed by Gibson assembly from various PCR-amplified fragments. For the construction of pUD801 these fragments were: the *ScTEF1p-kanR-ScTDH1t* G418 resistance marker cassette from pUD740 (primers 12100 + 4377), upstream (primers 14385 + 14386) and downstream (primers 14387 + 14388) homologous recombination sequences of the *NDH2-1* locus from CBS11895 genomic DNA and the plasmid backbone from pUD602 (primers 12103 + 12104). For pUD802 these fragments were: the *AgTEF1p-hph-AgTEF1t* hygromycin resistance marker cassette from pUDP002 (primers 11065 + 11133), upstream (primers 14389 + 14390) and downstream (primers 14391 + 14392) homologous recombination sequences of the *NDH2-2* locus from CBS11895 genomic DNA and the plasmid backbone from pUD602 (primers 12103 + 12104). For pUD803 these fragments were: the *AaTEF1* promoter from pUDP002 (primers 10426 + 14383), the *NatR* nourseothricin resistance marker ORF from pYTK078 (primers 14381 + 14382), the *ScPHO5* terminator from pUDP002 (primers 10427 + 14384), upstream (primers 14393 + 14394) and downstream (primers 14395 + 14396) homologous recombination sequences of the *NDH2-3* locus from CBS11895 genomic DNA and the plasmid backbone from pUD602 (primers 12103 + 12104). For pUD1035 these fragments were: the *AgTEF1p-pat-AgTEF1t* phosphinothricin resistance marker cassette from pAG31 (primers 3242 + 8439), upstream (primers 14803 + 14804) and downstream (primers 14805 + 14806) homologous recombination sequences for *GUT2* from CBS11895 genomic DNA and the plasmid backbone from pUD602 (primers 12103 + 12104). For pUD1036 these fragments were: the *AgTEF1p-pat-AgTEF1t* phosphinothricin resistance marker cassette from pAG31 (primers 3242 + 8439), upstream (primers 14807 + 14808) and downstream (primers 14809 + 14810) homologous recombination sequences for *NUBM* from CBS11895 genomic DNA and the plasmid backbone from pUD602 (primers 12103 + 12104). Correct insertion and presence of the homology-flanked marker cassettes in the constructed plasmids was verified by restriction digest and diagnostic PCR using PvuI + NdeI and primer sets 2908 + 12616 and 1642 + 3983 (pUD801), PvuI + PsiI and primer sets 2457 + 12616 and 1642 + 1781 (pUD802), PvuI + KpnI and primer sets 10459 + 12616 and 1642 + 10458 (pUD803) and PvuI + NcoI and primer sets 1409 + 12616 and 1642 + 4662 (pUD1035 & pUD1036).

Plasmids pTrc99A-NDH2-1, pTrc99A-NDH2-2 and pTrc99A-NDH2-3 were constructed by restriction/ligation cloning. The ORFs encoding *NDH2-1* (HPODL_02792), *NDH2-2* (HPODL_00256) and *NDH2-3* (HPODL_02018) were PCR-amplified from CBS11895 genomic DNA using primer sets 14929 + 14931, 16075 + 16076 and 16077 + 16078, respectively. Amplification using these primer sets added a 5' NcoI site, a GS-flanked 6-HIS tag ('GSHHHHHHGS') directly after the start codon and

a 3' XmaI site to all three ORFs. Furthermore, amplification of *NDH2-1* omitted the first 24 amino acids after the start codon, as they could be unambiguously identified as mitochondrial targeting sequence by MitoFates (Fukasawa *et al.* 2015). PCR amplicons were then digested with NcoI and XmaI and cloned into NcoI/XmaI-digested pTrc99A using T4 DNA ligase (New England Biolabs, Ipswich, MA, USA).

Yeast strain construction. *O. parapolymorpha* strains were transformed via electroporation of freshly prepared electrocompetent cells as described previously (Juergens *et al.* 2018b). Depending on the selection marker, mutants were selected on solid YPD medium supplemented with 200 µg mL^{-1} G418, 300 µg mL^{-1} hygromycin B or 100 µg mL^{-1} nourseothricin, or on solid synthetic medium (SM) supplemented with 20 g L^{-1} glucose and 200 µg mL^{-1} bialaphos (SanBio, Uden, The Netherlands). SM was prepared according to Verduyn *et al.* (1992) and autoclaved at 120°C for 20 min. Glucose and vitamins (Verduyn *et al.* 1992) were prepared separately and filter-sterilized (vitamins) or heat-sterilized at 110°C for 20 min (glucose).

Strains IMD003, IMD004 and IMD005 with disrupted versions of genes *NDH2-1* (HPODL_02792), *NDH2-2* (HPODL_00256) and *NDH2-3* (HPODL_02018), respectively, were constructed using the pUDP CRISPR/Cas9 system described previously (Juergens *et al.* 2018b). Wild type strain CBS11895 was transformed with pUDP019, pUDP020 or pUDP021 targeting *NDH2-1* (after base part 269 out of 1614), *NDH2-2* (after base pair 290 out of 1671) and *NDH2-3* (after base pair 515 out of 2097), respectively, and subjected to the prolonged liquid incubation protocol as described previously for deletion of *OpADE2* and *OpKU80* (Juergens *et al.* 2018b). Randomly picked colonies were then subjected to PCR amplification of the *NDH2-1*, *NDH2-2* and *NDH2-3* locus using primer sets 10742 + 10743, 10744 + 10745 and 10746 + 10747, respectively, followed by Sanger sequencing (Baseclear, Leiden, The Netherlands) to identify mutant transformants harboring a frame-shifting indel at the respective gRNA target sites. Three mutants with either a deletion of base pairs 226-229 of *NDH2-1*, an additional thymine nucleotide between position 289 and 290 of *NDH2-2* or a deletion of base pairs 505-515 of *NDH2-3* were identified, restreaked three times subsequently on non-selective YPD medium to remove the pUDP plasmids, and renamed IMD003, IMD004 and IMD005, respectively.

Strains IMX1945, IMX1978, IMX2017, IMX2197 and IMX2167 were constructed using a split-marker deletion approach (Fairhead *et al.* 1996), with ~480 bp of internal (marker recombination) and ~480 bp of external (genome recombination) homology. To preserve the promoter and terminator sequences of neighboring genes and limit interference of their expression, a minimum of 800 bp preceding and 300 bp succeeding adjacent ORFs were kept unaffected by the deletions. IMX1945 was constructed from wild type *O. parapolymorpha* strain CBS11895 by deletion of *NDH2-2* with a *hyg*

resistance cassette using pUD802, IMX1978 was constructed from IMX1945 by additional deletion of *NDH2-3* with a *NatR* resistance cassette using pUD803, IMX2017 was constructed from IMX1978 by additional deletion of *NDH2-1* with a *kanR* resistance cassette using pUD801, IMX2197 was constructed from IMX1978 by additional deletion of *NUBM* (HPODL_04625) (Juergens *et al.* 2020a) with a *pat* resistance cassette using pUD1036, and IMX2167 was constructed from IMX2017 by additional deletion of *GUT2* (HPODL_00581) with a *pat* resistance cassette using pUD1035. For the split-marker deletion of *NDH2-1*, *NDH2-2*, *NDH2-3*, *NUBM* and *GUT2*, the two overlapping fragments for transformation were PCR-amplified from pUD801 using primer sets 6816 + 14397 and 12565 + 14398, from pUD802 using primer sets 14399 + 14400 and 14401 + 14402, from pUD803 using primer sets 14403 + 14404 and 14405 + 14406, from pUD1036 using primer sets 15884 + 15885 and 15886 + 15887, and from pUD1035 using primer sets 14811 + 15885 and 14812 + 15886, respectively. Prior to transformation, the amplified fragments were gel-purified and, in case DNA amounts were too low for transformation, used as template for another PCR amplification using the same primers followed by PCR purification. For each transformation, a total of ~1 µg purified DNA (both fragments equimolar) in a maximum volume of 4 µL was transformed to 40 µL of fresh electrocompetent cells as described above, with the exception that after electroporation the cell suspensions were recovered in 1 mL YPD for 3 h at 30°C before plating onto selective medium. Additionally, after YPD recovery, cells transformed with the *pat* resistance marker were washed once by centrifugation and resuspension in sterile demineralized water before selective plating. Selection plates were typically incubated for 3 days at 30°C before assessment of the correct replacement of the target genes with the resistance markers via diagnostic PCR using primer sets 14465 + 14466, 14465 + 4047 and 2653 + 14466 for *ΔNDH2-1::kanR*, primer sets 14467 + 14468, 14467 + 7864 and 8411 + 14468 for *ΔNDH2-2::hph*, primer sets 14469 + 14470, 14469 + 11197 and 11202 + 14470 for *ΔNDH2-3::NatR*, primer sets 15630 + 15631, 15630 + 15885 and 15886 + 15631 for *ΔNUBM::pat* and primer sets 15628 + 15629, 15628 + 15885 and 15886 + 15629 for *ΔGUT2::pat*. Single colonies that contained the desired genotype(s) were restreaked once on selective medium, followed by two restreaks on non-selective YPD medium before stocking.

Molecular biology. PCR amplification for cloning and construction was performed with Phusion High Fidelity Polymerase (Thermo Fisher Scientific) using PAGE-purified oligonucleotide primers (Sigma-Aldrich, St. Louis, MO, USA) according to manufacturer's recommendations, with the exception that a final primer concentration of 0.2 µM was used. Diagnostic PCR was done using DreamTaq polymerase (Thermo Fisher Scientific) and desalted primers (Sigma-Aldrich). The primers used in this study are shown in **Table S5.4**. Genomic DNA of yeast colonies was isolated using the

LiAc-sodium dodecyl sulfate method (Lõoke *et al.* 2011) or the YeaStar Genomic DNA kit (Zymo Research, Irvine, CA, USA). DNA fragments obtained by PCR were separated by gel electrophoresis. Gel-purification was carried out using the Zymoclean Gel DNA Recovery Kit (Zymo Research). PCR purification was performed using the GenElute PCR Clean-Up Kit (Sigma-Aldrich). Gibson assembly was done using the NEBuilder HiFi DNA Assembly Master Mix (New England Biolabs) with purified DNA fragments according to manufacturer's recommendations, with the exception that reaction volume was down-scaled to 5-10 µL. DNA fragments that were PCR-amplified from a template harboring the same bacterial resistance marker as the construct to be Gibson-assembled were subjected to DpnI treatment prior to PCR cleanup. Restriction digest was performed using FastDigest enzymes (Thermo Fisher Scientific) or High Fidelity (HF) restriction endonucleases (New England Biolabs, Ipswich, MA, USA) according to the manufacturer's instructions. *E. coli* strains XL1-blue and DH5α were used for plasmid transformation, amplification and storage. Plasmid isolation from *E. coli* was done using the GenElute Plasmid Miniprep Kit (Sigma-Aldrich) or the Monarch Plasmid Miniprep Kit (New England Biolabs).

Multiple sequence alignment and domain prediction. Alignment of NDH2 protein sequences was done using MUSCLE (https://www.ebi.ac.uk/Tools/msa/muscle/) (Edgar 2004) and visualized using Jalview (Waterhouse *et al.* 2009). Orientation and substrate specificity of other fungal NDH2s were taken from: *K. lactis* (Tarrio *et al.* 2006, Tarrio *et al.* 2005), *N. crassa* (Carneiro *et al.* 2004, Carneiro *et al.* 2007, Duarte *et al.* 2003, Melo *et al.* 2001), *S. cerevisiae* (de Vries and Marres 1987, Luttik *et al.* 1998, Small and McAlister-Henn 1998, Van Urk *et al.* 1989), and *Y. lipolytica* (Kerscher *et al.* 1999). Prediction of the putative EF-hand calcium binding domain in NDH2 sequences was done by Motif Scan (https://myhits.isb-sib.ch/cgi-bin/motif_scan) using PROSITE profiles (Sigrist *et al.* 2010).

Shake-flask cultivation. Shake-flask growth experiments were performed with synthetic medium with urea as nitrogen source (Luttik *et al.* 2000), set to an initial pH of 5.0 with KOH. Cultures were grown in 500 mL round-bottom shake flasks filled with 50 mL medium. Cultures were grown with 2 g L^{-1} glucose as sole carbon source and were inoculated with mid-exponential precultures (washed once with sterile demineralized water) to an initial OD_{660} of 0.3. Precultures were grown under the same conditions and in the same medium, but with an initial glucose concentration of 5 g L^{-1}. Shake flasks were continuously shaken during sampling to prevent oxygen limitation. Physiological parameters were calculated from at least 5 samples taken during the exponential growth phase. Calculated ethanol yields were not corrected for evaporation.

Chemostat cultivation. Chemostat cultivation was performed as described previously (Juergens *et al.* 2020a) using SM with the addition of 0.15 g L^{-1} Pluronic 6100 PE antifoaming agent (BASF, Ludwigshafen, Germany) and glucose (7.5 or 9 g L^{-1}) as sole carbon source. SM was prepared according to Verduyn *et al.* (1992) as described above. Bioreactors were inoculated with exponentially growing shake flask cultures (SM with 20 g L^{-1} glucose). Chemostat cultivation was performed in 2-L benchtop bioreactors (Applikon, Delft, The Netherlands) with a working volume of 1.0 L which was maintained by an electrical level sensor that controlled the effluent pump. The dilution rate was set by maintaining a constant medium inflow rate. Cultures were sparged with dried, compressed air (0.5 vvm) and stirred at 800 rpm. Temperature was maintained at 30°C and pH was controlled at 5.0 by automatic addition of 2 M KOH by an EZcontroller (Applikon). The exhaust gas was cooled with a condenser (2°C) and dried with a Perma Pure Dryer (Inacom Instruments, Veenendaal, the Netherlands) prior to online analysis of carbon dioxide and oxygen with a Rosemount NGA 2000 Analyzer (Emerson, St. Louis, MO, USA). Cultures were assumed to have reached steady state when, after a minimum of 5 volume changes, the oxygen-consumption rate, carbon-dioxide production rate and biomass concentration changed by less than 3% over two consecutive volume changes.

Analytical methods. Optical density (OD) of yeast cultures was measured at 660 nm on a Jenway 7200 spectrophotometer (Jenway, Staffordshire, UK). OD of bacterial cultures was measured at 600 nm on Ultrospec 2100 pro (Amersham, Little Chalfont, UK). For biomass dry weight determination of yeast cultures, exactly 10 mL of culture broth was filtered over pre-dried and pre-weighed membrane filters (0.45 µm, Pall corporation, Ann Arbor, MI, USA), which were washed with demineralized water, dried in a microwave oven at 350 W for 20 min and weighed immediately (Postma *et al.* 1989). Samples were diluted with demineralized water prior to filtration to obtain a biomass dry weight concentration of approximately 2 g L^{-1}. The exact dilution was calculated by weighing the amount of sample and diluent and assuming a density of 1 g mL^{-1} for both fractions. Concentrations of extracellular metabolites and putative alcoholic contaminants of NAD(P)H substrates were analyzed by high-performance liquid chromatography (HPLC) on an Agilent 1100 HPLC (Agilent Technologies, Santa Clara, CA USA) with an Aminex HPX-87H ion-exchange column (BioRad, Veenendaal, The Netherlands) operated at 60°C with 5 mM H_2SO_4 as mobile phase at a flow rate of 0.6 mL min^{-1}. For the determination of extracellular metabolites, 1 mL aliquots of culture broth were centrifuged for 3 min at 20,000 g and the supernatant was used for analysis. Protein concentrations of mitochondrial preparations were estimated by the Lowry method (Lowry *et al.* 1951), using dried bovine serum albumin (BSA, fatty acid-

free, Sigma-Aldrich) as standard. Where necessary, protein determinations were corrected for BSA present in the mitochondrial preparations. Protein concentrations of *E. coli* membrane fractions were determined using a bicinchoninic acid (Smith *et al.* 1985) protein assay kit (Sigma-Aldrich) with BSA (Interchim, Montlucon, France) as standard.

Isolation of mitochondrial fractions. Mitochondria were isolated from glucose-limited, aerobic chemostat cultures (D = 0.1 h^{-1}) according to a procedure similar as described for *S. cerevisiae* (Luttik *et al.* 1998), based on the mild osmotic lysis method developed for *Candida utilis* (Bruinenberg *et al.* 1985a). Biomass (1.5 g dry weight) was harvested by centrifugation at 3000 g for 4 min. The pellet was then resuspended by vortexing in 30 mL of Tris buffer (100 mM) containing 10 mM dithiothreitol (final buffer pH of 9.3) and incubated at 30°C for 10 min. Afterwards, the cells were washed twice with 30 mL buffer A (25 mM potassium phosphate, 2 M sorbitol, 1 mM MgCl2, 1 mM EDTA, pH 7.5) by centrifugation (4000 g, 4 min) and resuspension (gentle vortexing). Then, cells were pelleted again by centrifugation (5000 g, 8 min) and resuspended (gentle vortexing) in a total volume of 40 mL buffer A. 3.06 mg of zymolyase (from *Arthrobacter luteus*, 20,000 U g^{-1}, AMS Biotechnology, Abingdon, UK) dissolved in 200 µL buffer A was added to the cell suspension, which was subsequently incubated at 30°C under gentle shaking for 60-90 min. Incubation time depended on the rate of spheroplast formation which was estimated based on the sensitivity to osmotic shock by 200-fold dilution in demineralized water as described previously (Bruinenberg *et al.* 1985a). Incubation was continued until osmotic resistance decreased to approx. 25% (see **Figure S5.3**). During the zymolyase treatment, release of glucose-6-phosphate dehydrogenase activity from compromised cells was measured as described previously (Bruinenberg *et al.* 1985a, Bruinenberg *et al.* 1983a) and typically did not exceed 5% compared to a sonicated sample. After zymolyase treatment, all subsequent steps were carried out on ice or in a cooled (4°C) centrifuge. Spheroplasts were washed twice with 35 mL buffer A by centrifugation (4400 g, 7 min) and resuspension (gentle shaking), followed by centrifugation (5000 g, 6 min) and resuspension (gentle shaking) in a total volume of 10 mL buffer A. Subsequently, 30 mL of buffer B (25 mM potassium phosphate, 0.2 M sorbitol, 1 mM $MgCl_2$, 1 mM EDTA, pH 7.5) was added dropwise to the spheroplast suspension over a timeframe of ~2-3 h, while it was slowly stirred with a magnetic stirrer bar. The spheroplast suspension was then subjected to 2 strokes in a cooled Potter-Elvehjem homogenizer (150 rpm, clearance 28 µm). After centrifugation (3000 g, 10 min), the supernatant was separated from intact cells and debris and spun again (12000 g, 10 min). The resulting pellet, containing the mitochondria, was resuspended in 2.5 mL of buffer C (25 mM potassium

phosphate, 0.65 M sorbitol, 1 mM $MgCl_2$, 1 mM EDTA, 1 mg mL^{-1} BSA (fatty acid-free, Sigma-Aldrich), pH 7.5) and kept on ice.

Oxygen-uptake studies with mitochondrial preparations. Substrate-dependent oxygen consumption rates of mitochondria were determined polarographically at 30°C with a Clark-type oxygen electrode. The assay mixture (4 mL) contained 25 mM potassium phosphate buffer (pH 7.0), 5 mM $MgCl_2$, and 0.65 M sorbitol. Reactions were started with ethanol (5 mM), methanol (5 mM), L-malate + pyruvate (both 5 mM, adjusted to pH 7.0 with KOH), 0.25 mM NADH (Prozomix, Haltwhistle, UK) or 0.75 mM NADPH (Oriental Yeast Co., Tokyo, Japan). While some commercial preparations of NADH and NADPH are contaminated with ethanol (Luttik *et al.* 1998, Overkamp *et al.* 2002, Patchett and Jones 1986), no ethanol (or methanol) was detected via HPLC analysis in freshly prepared, concentrated (100 mM) solutions of the NADH and NADPH used in this study (detection limit: ethanol 1 mM; methanol 5 mM). Oxygen uptake rates were calculated based on a dissolved oxygen concentration of 236 µM in air-saturated water at 30°C. Respiratory control values were determined by adding 0.25 mM ADP (Chance and Williams 1956). For tests with rotenone (50 µM), a concentrated stock solution (20 mM in DMSO) was freshly prepared directly before the respective assays and kept at room temperature. Mitochondria were pre-incubated in the presence of rotenone for 5 min at assay conditions prior to substrate addition. Preincubation with equivalent amounts of DMSO without rotenone did not measurably affect oxygen uptake rates. Tests with KCN (1 mM) were conducted with a concentrated stock solution (200 mM, in 100 mM NaOH), which was added to mitochondria during ADP-stimulated respiration (state III). Addition of equivalent amounts of NaOH without KCN affected oxygen uptake rates by less than 15%.

Overexpression of *O. parapolymorpha* NDH2s in *E. coli*. Overexpression and purification strategy were based on previous literature (Godoy-Hernandez *et al.* 2019). *E. coli* BL21(DE3) cells were transformed with plasmids pTrc99A, pTrc99A-NDH2-1, pTrc99A-NDH2-2, and pTrc99A-NDH2-3. Correct transformants were pre-cultured (37°C, 180 rpm) for 16 h in lysogeny broth with 100 µg mL^{-1} ampicillin. These precultures were used to inoculate 500 mL apple flasks with 100 mL 2xYT medium (16 g L^{-1} tryptone, 10 g L^{-1} yeast extract, 5 g L^{-1} NaCl) supplemented with 20 g L^{-1} glucose and 100 µg mL^{-1} ampicillin and grown under the same conditions. Once cultures had reached an OD_{600} of 0.5, NDH2 overexpression was induced using 1 mM isopropyl β-D-1-thiogalactopyranoside (IPTG, Sigma Aldrich), followed by growth for an additional 5 h at 30°C and 180 rpm. Afterwards, cells were harvested by centrifugation (7000 g, 10 min) and washed with 25 mL buffer W1 (50 mM Tris-HCl, pH 8.0, 2 mM $MgCl_2$). Cell pellets were then resuspended in buffer W1 (at 4 mL per g wet weight),

containing 0.1 mM PMSF (Sigma-Aldrich) and 0.1 mg mL^{-1} bovine pancreatic DNase (New England Biolabs), and incubated 10 min at room temperature. Cells were then disrupted using a cell disruptor (Constant Systems Ltd., Daventry, UK) at 1.38 kbar. Unbroken cells and cell debris were removed by centrifugation (10,000 g, 10 min). The membrane fraction was isolated from the cell lysate by ultracentrifugation (180,000 g, 45 min, 4°C), and the resulting membrane pellet was resuspended in buffer W1 and stored at -80°C.

***In vitro* NAD(P)H dehydrogenase activity tests with NDH2s.** NAD(P)H:quinone oxidoreductase activity was measured using a spectrophotometric assay similar to as described previously (Godoy-Hernandez *et al.* 2019). Activity was monitored spectrophotometrically using a modified Cary 60 UV/Vis Spectrophotometer (Agilent Technologies), following the oxidation of NADH or NADPH at 340 nm in the presence of ubiquinone-1 (UQ-1, Sigma-Aldrich) at 37°C. Membrane preparations (10 µg protein mL^{-1}) and UQ-1 (100 mM) were added to pre-warmed reaction buffer (final volume 2 mL) in a 1 cm path length cuvette and incubated for 30 s. Depending on the pH, the reaction buffer consisted of i) 50 mM Tris-HCl (pH 8.0), 150 mM NaCl, ii) 25 mM MES + 25 mM MOPS (pH 7.4), 150 mM NaCl or iii) 25 mM MES + 25 mM MOPS (pH 5.5), 150 mM NaCl. NADH or NADPH (200 µM at pH 8.0 and 7.4, 100 µM at pH 5.5) were added to the mixture to initiate the reaction. An extinction coefficient of 6.3 mM^{-1} cm^{-1} was used to calculate NAD(P)H concentration. For tests with calcium, 5 mM $CaCl_2$ was added to the assay after 1 min of reaction time had elapsed.

Acknowledgements

The authors thank Veronica Gast for help with plasmid and strain construction, Lisan Broekman for performing part of the shake flask characterization, Marijke Luttik for providing technical expertise and assistance regarding isolation and measurement of mitochondria, Erik de Hulster for assistance with chemostat cultivation and Jack Pronk for helpful discussions and feedback on the manuscript.

Supplementary material

Figure S5.1: Section of MUSCLE sequence alignment of fungal/yeast NDH2 sequences, visualized by Jalview using Clustalx residue colors. Black boxes denote (left) conserved GxGxxGxE motif of second dinucleotide (substrate) binding domain and (right) residue affecting NADH/NADPH substrate specificity. Orientation, either I (internal) or E (external) and substrate utilization, NADH, NADPH or NAD(P)H (= both) are depicted for characterized fungal NDH2s. *Kl*, *Kluyveromyces lactis*, *Nc*, *Neurospora crassa*, *Op*, *Ogataea parapolymorpha*, *Sc*, *Saccharomyces cerevisiae*, *Yl*, *Yarrowia lipolytica*.

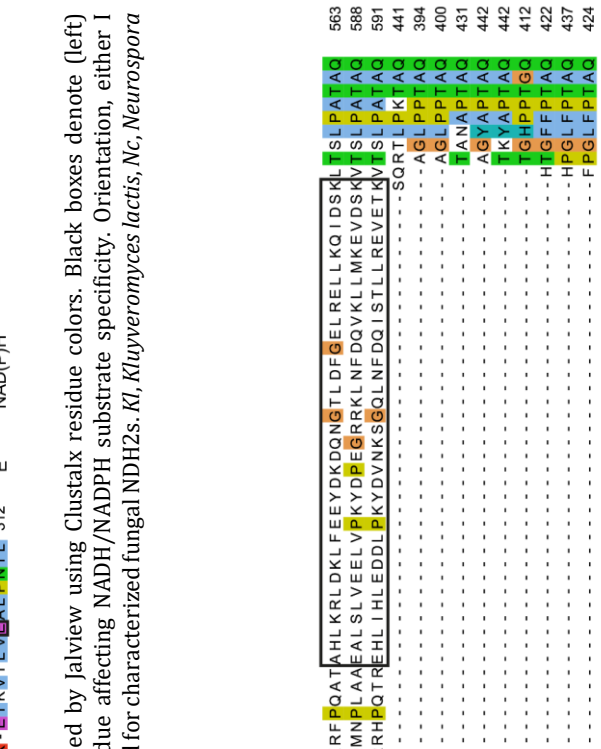


Figure S5.2: Section of MUSCLE multiple sequence alignment of fungal/yeast NDH2 sequences, visualized by Jalview using Clustalx residue colors. The black box denotes a (putative) EF-hand calcium binding domain as predicted by Motif Scan. Kl, Kluyveromyces lactis, Nc, Neurospora crassa, Op, *Ogataea parapolymorpha*, *Sc*, *Saccharomyces cerevisiae*, *Yl*, *Yarrowia lipolytica*.

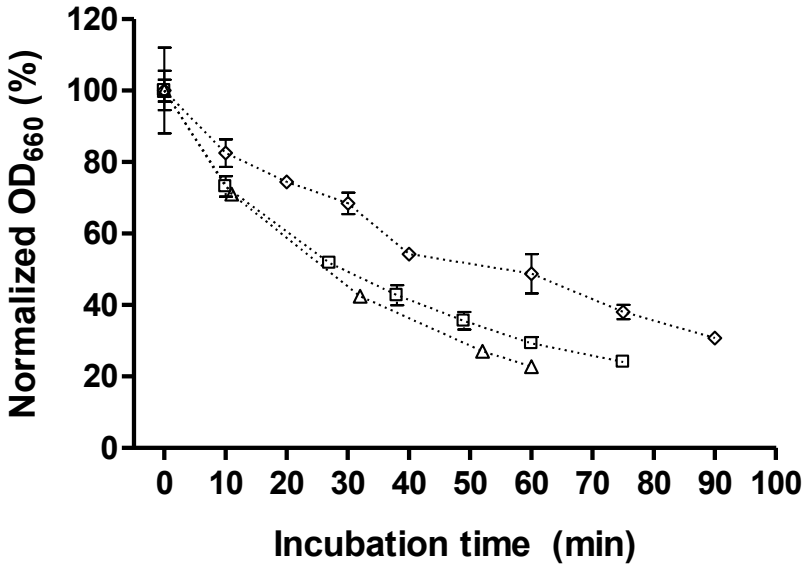


Figure S5.3: Proportion of osmotically insensitive cells of *O. parapolymorpha* strains CBS11895 (triangles), IMX2017 (diamonds) and IMX2197 (squares) during zymolyase treatment as determined by dilution in demineralized water. Data is presented as mean ± standard deviation from at least 2 independent replicates, and was normalized to initial OD for each strain individually.

Table S5.1: NAD(P)H oxidation (µmol (mg protein)$^{-1}$ min^{-1}) by *E. coli* membranes isolated from strains overexpressing individual *O. parapolymorpha* NDH2s. Control measurements were done with membranes isolated from a strain carrying an empty overexpression plasmid (pTrc99A). Assays were performed with a membrane protein concentration of 10 µg mL^{-1} at 37°C, with 200 µM (pH 7.4 and 8) or 100 µM (pH 5.5) NAD(P)H and 100 µM ubiquinone-1. For tests with calcium, 5 mM CaCl was used. Data is presented as mean ± standard deviation of at least duplicate measurements. [A]Single measurement.

	NADH			NADPH		
	pH 5.5	pH 7.4	pH 8.0	pH 5.5	pH 7.4	pH 8.0
Control pTrc99A	1.64 ± 0.34	1.64 ± 0.34	1.49 ± 0.11	0.03 ± 0.04	0.02 ± 0.03	0.18 ± 0.01
Ndh2-1 pTrc99A-NDH2-1	4.47 ± 0.05	5.27 ± 0.82	4.58 ± 0.25	< 0.01	0.03 ± 0.05	0.05 ± 0.05
Ndh2-2 pTrc99A-NDH2-2	3.45 ± 0.13	3.95 ± 0.85	3.24 ± 0.94	0.06 ± 0.08	0.02 ± 0.04	0.09 ± 0.09
Ndh2-3 pTrc99A-NDH2-3	3.77 ± 0.05	4.34 ± 0.50	4.70 ± 0.02	0.21 ± 0.16	0.01 ± 0.02	0.03 ± 0.04
Control + Ca^{2+} pTrc99A	1.43[A]	1.64 ± 0.12	1.64 ± 0.09	0.07 ± 0.05	< 0.01	0.09 ± 0.02
Ndh2-3 + Ca^{2+} pTrc99A-NDH2-3	3.08 ± 0.25	4.15 ± 0.69	3.57 ± 0.09	0.17 ± 0.16	< 0.01	< 0.01

Table S5.2: Physiology of *Ogataea parapolymorpha* strain IMX2167 in aerobic, glucose-limited chemostat cultures grown at a dilution rate of 0.1 h^{-1} at 30°C and pH 5. Data are presented as mean ± mean absolute deviation from two independent replicates. Carbon recoveries were calculated based on a biomass carbon content of 48% (w/w). Symbols: $Y_{X/S}$ and $Y_{X/O2}$ = yield of biomass dry weight on glucose and oxygen, respectively; RQ = respiratory quotient; $q_{Glucose}$, q_{CO2} and q_{O2} represent biomass-specific uptake/production rates of glucose, CO_2 and O_2, respectively; C_X represents biomass dry weight concentration.

Strain	**IMX2167** *Δndh2-1 Δndh2-2 Δndh2-3 Δgut2*
Actual dilution rate (h^{-1})	0.10 ± 0.00
Reservoir glucose (g L^{-1})	7.61 ± 0.01
$Y_{X/S}$ (g biomass [g glucose]$^{-1}$)	0.52 ± 0.00
$Y_{X/O2}$ (g biomass [g O_2]$^{-1}$)	1.44 ± 0.00
RQ (-)	1.05 ± 0.00
$q_{Glucose}$ (mmol [g biomass]$^{-1}$ h^{-1})	-1.04 ± 0.00
q_{CO2} (mmol [g biomass]$^{-1}$ h^{-1})	2.25 ± 0.01
q_{O2} (mmol [g biomass]$^{-1}$ h^{-1})	-2.14 ± 0.00
C_X (g biomass L^{-1})	3.97 ± 0.00
Carbon recovery (%)	98.7 ± 0.3

Table S5.3: *S. cerevisiae* proteins (proposed to be) involved in ethanol-acetaldehyde, malate-oxaloacetate and malate-aspartate shuttle NADH shuttles, and corresponding ortholog proteins in *O. parapolymorpha*. Orthologs were identified using blastp (https://blast.ncbi.nlm.nih.gov/). *HPODL_03666 and HPODL_02528 have been experimentally verified as cytosolic and mitochondrial alcohol dehydrogenases in *O. parapolymorpha*, respectively (Suwannarangsee *et al.* 2012, Suwannarangsee *et al.* 2010).

***S. cerevisiae* protein**	***Annotation***	***O. parapolymorpha* ortholog**	***E-value (blastp)***
Adh1	Cytosolic alcohol dehydrogenase	HPODL_03666*	0.0
Adh3	Mitochondrial alcohol dehydrogenase	HPODL_02528*	0.0
Mdh1	Mitochondrial malate dehydrogenase	HPODL_00710	$2*10^{-174}$
Mdh2	Cytosolic malate dehydrogenase	HPODL_02341	$2*10^{-83}$
Aat1	Mitochondrial aspartate aminotransferase	HPODL_01819	$2*10^{-91}$
Aat2	Cytosolic aspartate aminotransferase	HPODL_02844	0.0
Dic1	Mitochondrial dicarboxylate carrier	HPODL_04016	$9*10^{-112}$
Oac1	Mitochondrial oxaloacetate carrier	HPODL_00744	$5*10^{-157}$
Odc1 / Odc2	Mitochondrial oxodicarboxylate carrier	HPODL_01147	$8*10^{-144}$ / $3*10^{-140}$
Agc1	Mitochondrial amino acid transporter	HPODL_03446	$5*10^{-122}$

Table S5.4: Primers used in this study.

Name	Sequence (5'-3')	Purpose
12097	TTAGAAAAACTCATCGAGCATC	Construction of pUD740
12098	ATGGGTAAGGAAAAGACTCAC	Construction of pUD740
12099	ATCAGCATCCATGTTGGAATTTAATCGCGGCCTCGAAACGTGAGTCTTT TCCTTACCCATTTTGTAATTAAAACTTAGATTAGATTGCTATGC	Construction of pUD740
12100	CCTTGCCAACAGGGAGTTC	Construction of pUD740, pUD801
12095	ACCTTAAGTGCATATGCCGTATAAGGGAAACTCAAAGAACTGGCATCG CAAAAATGAAAACGTTCAGGGTAATATATTTTAAC	Construction of pUD740
12096	ATTGATAATCCTGATATGAATAAATTGCAGTTTCATTTGATGCTCGAT GAGTTTTTCTAAATAAAGCAATCTTGATGAGG	Construction of pUD740
10426	AATCTATAATCAGTCCATAGTCAACAAGAGCC	Construction of pUD740, pUD803
10427	TTTTCATTTTTGCGATGCCAGTTCTTTG	Construction of pUD740, pUD803
12093	TAATACGACTCACTATAGGGCGAATTGGCGGAAGGCCGTCAAGGCCGC ATCAACGAGCTCTGTGCCTAGGATTATGTCCTG	Construction of pUD740
12094	GAGATCACTAAACGATATCAACTTTGAGGGCTCTTGTTGACTATGGAC TGATTATAGATTATAGTCTGATACTCAGAAAGATCGAAG	Construction of pUD740
12101	AGGCTGTCAATAATTTCGTTTTGAGCCTCCATGTCTCTGAAGAACTCCC TGTTGGCAAGGATCGTGCTCAAAGCCTACAAG	Construction of pUD740
12102	CAGGTTTCCCGACTGGAAAGCGGGCAGTGAGCGGAAGGCCCATGAGGC CCAGAGGGTACCCAAATTATAGTACTTGTTTAGCTGCC	Construction of pUD740
12103	GAGCTCGTTGATGCGGC	Construction of pUD740, pUD801, pUD802, pUD803, pUD1035, pUD1036
12104	GGTACCCTCTGGGCCTC	Construction of pUD740, pUD801, pUD802, pUD803, pUD1035, pUD1036
4377	CAGTATTAGTCGCCGCTTAG	Construction of pUD801
14385	CAGGTTTCCCGACTGGAAAGCGGGCAGTGAGCGGAAGGCCCATGAGGC CCAGAGGGTACCTTGAGCTGGAGCCCAGCG	Construction of pUD801
14386	AGGCTGTCAATAATTTCGTTTTGAGCCTCCATGTCTCTGAAGAACTCCC TGTTGGCAAGGAATCTAAGATAGTAAAATGTTCAGGTGGAAGTTTG	Construction of pUD801
14387	ATATGCACTTAAGGTTTGATTTTGCTAATTAAAGGCGTGCCTAAGCGG CGACTAATACTGCGCTTTGGCCCAGATCAAGCC	Construction of pUD801
14388	TAATACGACTCACTATAGGGCGAATTGGCGGAAGGCCGTCAAGGCCGC ATCAACGAGCTCGGACGCACCCCTGGTTTATGC	Construction of pUD801
11065	ACTATATGTGAAGGCATGGCTATGG	Construction of pUD802
11133	GTTGAACATTCTTAGGCTGGTC	Construction of pUD802
14389	CAGGTTTCCCGACTGGAAAGCGGGCAGTGAGCGGAAGGCCCATGAGGC CCAGAGGGTACCTATCCCTAGAATTGAGGCTAATTCGGCTTC	Construction of pUD802
14390	GTGCCTATTGATGATCTGGCGGAATGTCTGCCGTGCCATAGCCATGCCT TCACATATAGTTATTTAGGATATTTCGAAAAAATACTGCCGACAAAC	Construction of pUD802
14391	CACCTTTCGAGAGGACGATGCCCGTGTCTAAATGATTCGACCAGCCTAA GAATGTTCAACCACTTGAAGCGTCTTGAGAGATACAAGAAC	Construction of pUD802
14392	TAATACGACTCACTATAGGGCGAATTGGCGGAAGGCCGTCAAGGCCGC ATCAACGAGCTCCGCTGGTGATTTCTCTGGGAGAC	Construction of pUD802
14383	AGCATCACCAGGGACTGATGTTCTGTATCTGTAGGCTGTATCATCTAA GGTAGTACCCATTGTTGATTATGTTTTTAAGAACTACTC	Construction of pUD803
14384	TATGATGGAACAGCATCTGACGGTGAACAGGCCTTGTATATGTCTATG CCTTGCCCTTAAGCGGTTTTGTATAACTAAATAATATTGG	Construction of pUD803
14381	ATGGGTACTACCTTAGATG	Construction of pUD803
14382	TTAAGGGCAAGGCATAGAC	Construction of pUD803
14393	CAGGTTTCCCGACTGGAAAGCGGGCAGTGAGCGGAAGGCCCATGAGGC CCAGAGGGTACCCCAAAGAGCCCAAATGTCTATATTTGAAAGG	Construction of pUD803
14394	GAGATCACTAAACGATATCAACTTTGAGGGCTCTTGTTGACTATGGAC TGATTATAGATTACTGTGCCACTGTGCGTACC	Construction of pUD803
14395	ACCTTAAGTGCATATGCCGTATAAGGGAAACTCAAAGAACTGGCATCG CAAAAATGAAAAACAGCTTCAACACTCTTCCGACAG	Construction of pUD803
14396	TAATACGACTCACTATAGGGCGAATTGGCGGAAGGCCGTCAAGGCCGC ATCAACGAGCTCTATTGGCAACAAACTCCGATCGGATATTG	Construction of pUD803
14803	CAGGTTTCCCGACTGGAAAGCGGGCAGTGAGCGGAAGGCCCATGAGGC CCAGAGGGTACCTACAAGAACTGTTTGCTGAGCTTGAG	Construction of pUD1035
14804	CCCTGAGCTGCGCACGTCAAGACTGTCAAGGAGGGTATTCTGGGCCTCC ATGTCGCTGGCGCATCCCTGGCGGAAAAATT	Construction of pUD1035
14805	TTAAGTGCGCAGAAAGTAATATCATGCGTCAATCGTATGTGAATGCTG GTCGCTATACTGTCCAACTTATGCCAACTTCGATCATC	Construction of pUD1035
14806	TAATACGACTCACTATAGGGCGAATTGGCGGAAGGCCGTCAAGGCCGC ATCAACGAGCTCGATAATATTTTGGGAGAGGAGGATTTGATATGG	Construction of pUD1035
3242	CAGTATAGCGACCAGCATTC	Construction of pUD1035, pUD1036
8439	GCCAGCGACATGGAGGCCCAGAATAC	Construction of pUD1035, pUD1036

Table S5.4 (cont.)

14807	CAGGTTTCCCGACTGGAAAGCGGGCAGTGAGCGGAAGGCCCATGAGGC CCAGAGGGTACCTTGGTACGGCCAGCTAAACG	Construction of pUD1036
14808	CCCTGAGCTGCGCACGTCAAGACTGTCAAGGAGGGTATTCTGGGCCTCC ATGTCGCTGGCTGCATTTCGGGAAACGTGAAC	Construction of pUD1036
14809	TTAAGTGCGCAGAAAGTAATATCATGCGTCAATCGTATGTGAATGCTG GTCGCTATACTGTGCCGCCCGCCTG	Construction of pUD1036
14810	TAATACGACTCACTATAGGGCGAATTGGCGGAAGGCCGTCAAGGCCGC ATCAACGAGCTCTTTTTCCAATCTATATCTTTATTTTTCATCAT	Construction of pUD1036
2908	GGATTGGGTGTGATGTAAGGATTCGC	Diagnostic PCR of pUD801
12612	CAACAACATCACTCCATCTC	Diagnostic PCR of pUD801
1642	TTTCCCAGTCACGACGTTG	Diagnostic PCR of pUD801, pUD802, pUD803, pUD1035, pUD1036
3983	AGACCGATACCAGGATCTTG	Diagnostic PCR of pUD801
2457	CGCACGTCAAGACTGTCAAG	Diagnostic PCR of pUD802
12616	CGAGTCAGTGAGCGAGGAAG	Diagnostic PCR of pUD801, pUD802, pUD803, pUD1035, pUD1036
1781	TACTCGCCGATAGTGGAAAC	Diagnostic PCR of pUD802
10459	TGGCGGTTACTCTAAAGACG	Diagnostic PCR of pUD803
10458	GAGGAGCCGGTCATTTATGG	Diagnostic PCR of pUD803
1409	TATTCTGGGCCTCCATGTCGCTGG	Diagnostic PCR of pUD1035, pUD1036
4662	GACATCATCTGCCCAGATGC	Diagnostic PCR of pUD1035, pUD1036
14929	CATGATCCATGGGCAGCCATCACCATCACCATCACGGCAGCTCTGCCCA GCGCAAGTC	Construction of pTrc99A-NDH2-1
14931	CATCAGCCCGGGCTACTCGTTAGTCAAGTCCCTACC	Construction of pTrc99A-NDH2-1
16075	CATGATCCATGGGCAGCCATCACCATCACCATCACGGCAGCCAACGACA ACTAGCGTCAGTGG	Construction of pTrc99A-NDH2-2
16076	CATCAGCCCGGGTTAGTCTTTGGAGCAGTCGCGAC	Construction of pTrc99A-NDH2-2
16077	CATGATCCATGGGCAGCCATCACCATCACCATCACGGCAGCTTCCCATC AATAATAAAGGTTGGCC	Construction of pTrc99A-NDH2-3
16078	CATCAGCCCGGGTTAAACAGTCAAGATATCCCTACC	Construction of pTrc99A-NDH2-3
10742	CCGTGCGCCTCAACACTATC	Diagnostic PCR of genomic *NDH2-1* disruption
10743	GACCTCCTTGGCGATGGATG	Diagnostic PCR of genomic *NDH2-1* disruption
10744	CGCCTGTTTGGACACTCCTC	Diagnostic PCR of genomic *NDH2-2* disruption
10745	TTCTGGCAACAGTCTGGATG	Diagnostic PCR of genomic *NDH2-2* disruption
10746	TGGTTGCACGGTAACTATGG	Diagnostic PCR of genomic *NDH2-3* disruption
10774	TTATGCCGTCTCAGGTCTCACAGCCAGTGTTCCTTAATCAAGGATACC	Diagnostic PCR of genomic *NDH2-3* disruption
6816	ATTCCGACTCGTCCAACATC	Amplification of left split-marker fragment from pUD801 for deletion of *NDH2-1*
14397	TTGAGCTGGAGCCCAG	Amplification of left split-marker fragment from pUD801 for deletion of *NDH2-1*
12565	AAAGGTAGCGTTGCCAATG	Amplification of right split-marker fragment from pUD801 for deletion of *NDH2-1*
14398	GGACGCACCCCTGG	Amplification of right split-marker fragment from pUD801 for deletion of *NDH2-1*
14399	TATCCCTAGAATTGAGGCTAATTCG	Amplification of left split-marker fragment from pUD802 for deletion of *NDH2-2*
14400	GCTCGAAGTAGCGCGTC	Amplification of left split-marker fragment from pUD802 for deletion of *NDH2-2*
14401	ATCTCCCGCCGTGC	Amplification of right split-marker fragment from pUD802 for deletion of *NDH2-2*
14402	CGCTGGTGATTTCTCTGGG	Amplification of right split-marker fragment from pUD802 for deletion of *NDH2-2*
14403	ATCATATAAAGCAGTATCTAAACCAC	Amplification of left split-marker fragment from pUD803 for deletion of *NDH2-3*
14404	CCAAAGAGCCCAAATGTCTA	Amplification of left split-marker fragment from pUD803 for deletion of *NDH2-3*
14405	TATTGGCAACAAACTCCG	Amplification of right split-marker fragment from pUD803 for deletion of *NDH2-3*
14406	CCACCGGTGATGGATTTAC	Amplification of right split-marker fragment from pUD803 for deletion of *NDH2-3*
14811	TACAAGAACTGTTTGCTGAGCTTGAG	Amplification of left split-marker fragment from pUD1035 for deletion of *GUT2*
14812	GATAATATTTTGGGAGAGGAGGATTTG	Amplification of right split-marker fragment from pUD1035 for deletion of *GUT2*

Table S5.4 (cont.)

15884	GTACGGCCAGCTAAACGC	Amplification of left split-marker fragment from pUD1036 for deletion of *NUBM*
15887	GTTCCGGTAACAAGAGGAC	Amplification of right split-marker fragment from pUD1036 for deletion of *NUBM*
15885	GTGCTTGTAGCCGGCTG	Amplification of left split-marker fragment from pUD1036 for deletion of *NUBM*; amplification of left split-marker fragment from pUD1035 for deletion of *GUT2*; diagnostic PCR of *ΔNUBM::pat* deletion; diagnostic PCR of *ΔGUT2::pat deletion*
15886	TCTGCGACATCGTCAATC	Amplification of right split-marker fragment from pUD1036 for deletion of *NUBM*; amplification of right split-marker fragment from pUD1035 for deletion of *GUT2*; diagnostic PCR of *ΔNUBM::pat* deletion; diagnostic PCR of *ΔGUT2::pat deletion*
14465	TACGTCGACTAACAAGGAGC	Diagnostic PCR of *ΔNDH2-1::kanR* deletion
14466	GCGGTCTCGACAGCTTTC	Diagnostic PCR of *ΔNDH2-1::kanR* deletion
4047	AACTCACCGAGGCAGTTCCATAG	Diagnostic PCR of *ΔNDH2-1::kanR* deletion
2653	GGCAATCAGGTGCGACAATC	Diagnostic PCR of *ΔNDH2-1::kanR* deletion
14467	CTAGTGAAATTCCACTCTATAAATCGATCAT	Diagnostic PCR of *ΔNDH2-2::hph* deletion
14468	AAAGTCGGAGAGTGTTTTAGACC	Diagnostic PCR of *ΔNDH2-2::hph* deletion
7864	TCTGGGCAGATGATGTCGAG	Diagnostic PCR of *ΔNDH2-2::hph* deletion
8411	CGTTGAATTGTCCCCACG	Diagnostic PCR of *ΔNDH2-2::hph* deletion
14469	ATGTACAGGAATCTATATTTTTAAACAGTCAAG	Diagnostic PCR of *ΔNDH2-3::NatR* deletion
14470	AACTATGGATCTTTGAAAAAAAAAGCAAAC	Diagnostic PCR of *ΔNDH2-3::NatR* deletion
11197	ATGGGTTCCTGACTGACTAC	Diagnostic PCR of *ΔNDH2-3::NatR* deletion
11202	CACACCGGAAATCAAGGCAT	Diagnostic PCR of *ΔNDH2-3::NatR* deletion
15628	GACAAGAACGACGCGCTTTCGG	Diagnostic PCR of *ΔGUT2::pat* deletion
15629	ACGGCCAATATACTGCACGACC	Diagnostic PCR of *ΔGUT2::pat* deletion
15630	ACTTGAATATGCGGCCTGCAGC	Diagnostic PCR of *ΔNUBM::pat* deletion
15631	CTGTCCGCCGTATCATCCTTCC	Diagnostic PCR of *ΔNUBM::pat* deletion

Outlook

The research described in this thesis provides a framework for future projects to investigate and improve respiratory energy coupling in *Ogataea* yeasts for bioprocess applications.

The research described in **Chapter 2** demonstrates the potential of CRISPR-Cas9-based genetic engineering in organisms in which homologous recombination (HR) is not the preferred DNA repair mechanism. The method developed in **Chapter 2** was used in **Chapters 4 and 5** to construct *O. parapolymorpha* strains with precise gene disruptions while leaving no scars in the genome, something that could not easily be achieved with conventional methods. Given the vast improvements of CRISPR-Cas9 over classical methods, new and enhanced variants are constantly being developed. For example, use of the Cas12a/Cpf1 nuclease, which is smaller and cheaper to engineer for different (multiple) genomic targets, has already been described for use in non-conventional yeasts (Yang *et al.* 2020). However, Cas9- and Cpf1-based DNA editing systems and similar approaches, such as argonaut nucleases, rely on creating a double strand break (DSB) and harnessing the innate DNA repair mechanisms of the targeted organism to integrate donor DNA. This drastically limits their potential in species with a low frequency of HR. Therefore, until entire genomes are routinely *de novo* synthesized to exact specifications, it might be more worthwhile for researchers to focus on increasing the frequency of HR in their species of choice rather than to exclusively focus on methods for generating the DSB. Alternatively, future research could focus on the development of novel genome editing approaches in these organisms that do not rely on DSBs and native DNA repair, such as prime editing (Anzalone *et al.* 2019).

The research described in **Chapters 4 and 5** systematically characterizes the physiological importance of different electron entry points into the respiratory chain of *O. parapolymorpha*, and demonstrates how this yeast selectively uses different NADH dehydrogenases based on substrate availability. Understanding respiratory chain architecture and how energy coupling is modulated is important for rational design of strains and conditions for industrial bioprocesses and can help to minimize the loss of carbon and additional costs typically associated with aerobic production processes. Despite the apparent exclusive use of respiratory Complex I by *O. parapolymorpha* under glucose-limited growth conditions and the demonstrated energetic benefit of respiratory Complex I compared to alternative NADH dehydrogenases, wild-type *O. parapolymorpha* exhibited a biomass yield on glucose of 0.5 g/g, similar to that of yeasts without Complex I such as *S. cerevisiae* (van Dijken *et al.* 2000). Both yeasts exhibit a similar biomass protein content of about 40% under these conditions

(Diderich *et al.* 1999), and as a result are expected to exhibit a similar energetic requirement for biomass formation. Therefore, other metabolic processes and factors likely consume the additional energy conserved by Complex I in *O. parapolymorpha*, which will require further research to be identified. The use of proton-coupled substrate uptake by *Ogataea* yeasts under these conditions (Karp and Alamae 1998) could contribute (a small part) of the observed energy gap. Therefore, replacement of the endogenous high affinity glucose transporter by a transporter that mediates facilitated diffusion has the potential to conserve additional ATP, with the positive side effect of a higher robustness to substrate concentration fluctuations typically present in large-scale substrate-limited fermentations. Another energy sink might be the higher maintenance energy requirements, for example due to higher protein turnover (Canelas *et al.* 2010, Hong *et al.* 2012), which could be another future engineering target.

Knowledge about the composition of the respiratory chain and the capacity of its components (**Chapters 4 and 5**) also offers opportunities for metabolic engineering of the respiratory chain itself in *O. parapolymorpha*. For example, inactivation of Complex I could be utilized to obtain strains with increased flux through central carbon metabolism under substrate-limited conditions while maintaining a fully respiratory phenotype. Alternatively, inactivation of the internal NADH dehydrogenase could make large amounts of redox equivalents for formation of reduced products available in the cytosol if overflow pathways such as alcoholic fermentation are removed. Finally, the use of strains with linearized respiratory chains that route all electrons through the efficient Complex I pathway, such as constructed in **Chapter 5**, might also provide benefits when co-feeding of cultures with auxiliary substrates is considered. A suitable substrate could be formate, which can be obtained with (green) electricity from carbon dioxide (Wang *et al.* 2015). Co-feeding aerobic yeast cultures with formate has the potential to significantly increase product yields on the main (plant-derived) carbon source, as cytosolic formate oxidation can provide NADH that can be respired to gain ATP. Respiration of this NADH by Complex I instead of external alternative NADH dehydrogenase would further increase the efficiency of this process by maximizing the ATP gain per NADH respired.

References

Abdel-Banat BM, Nonklang S, Hoshida H, Akada R. Random and targeted gene integrations through the control of non-homologous end joining in the yeast *Kluyveromyces marxianus*. *Yeast* 2010;**27**: 29-39.

Altenhoff AM, Glover NM, Train CM, Kaleb K, Warwick Vesztrocy A, Dylus D, . . . Dessimoz C. The OMA orthology database in 2018: retrieving evolutionary relationships among all domains of life through richer web and programmatic interfaces. *Nucleic Acids Res* 2018;**46**: D477-D85.

Amann E, Ochs B, Abel KJ. Tightly regulated tac promoter vectors useful for the expression of unfused and fused proteins in *Escherichia coli*. *Gene* 1988;**69**: 301-15.

Anders S, Pyl PT, Huber W. HTSeq--a Python framework to work with high-throughput sequencing data. *Bioinformatics* 2015;**31**: 166-9.

Antos-Krzeminska N, Jarmuszkiewicz W. Alternative Type II NAD(P)H Dehydrogenases in the Mitochondria of Protists and Fungi. *Protist* 2019;**170**: 21-37.

Anzalone AV, Randolph PB, Davis JR, Sousa AA, Koblan LW, Levy JM, . . . Liu DR. Search-and-replace genome editing without double-strand breaks or donor DNA. *Nature* 2019;**576**: 149-57.

Arber W, Dussoix D. Host specificity of DNA produced by *Escherichia coli*. I. Host controlled modification of bacteriophage lambda. *J Mol Biol* 1962;**5**: 18-36.

Arber W, Linn S. DNA modification and restriction. *Annu Rev Biochem* 1969;**38**: 467-500.

Arbige M, Chesbro WR. Very Slow Growth of *Bacillus polymyxa*: Stringent Response and Maintenance Energy. *Arch Microbiol* 1982;**132**: 338-44.

Ata O, Rebnegger C, Tatto NE, Valli M, Mairinger T, Hann S, . . . Mattanovich D. A single Gal4-like transcription factor activates the Crabtree effect in *Komagataella phaffii*. *Nat Commun* 2018;**9**: 4911.

Avila J, Perez MD, Brito N, Gonzalez C, Siverio JM. Cloning and disruption of the *YNR1* gene encoding the nitrate reductase apoenzyme of the yeast *Hansenula polymorpha*. *FEBS Lett* 1995;**366**: 137-42.

Baerends RJ, de Hulster E, Geertman JM, Daran JM, van Maris AJ, Veenhuis M, . . . Pronk JT. Engineering and analysis of a *Saccharomyces cerevisiae* strain that uses formaldehyde as an auxiliary substrate. *Appl Environ Microbiol* 2008;**74**: 3182-8.

Baeshen NA, Baeshen MN, Sheikh A, Bora RS, Ahmed MM, Ramadan HA, . . . Redwan EM. Cell factories for insulin production. *Microb Cell Fact* 2014;**13**: 141.

Baker M. 1,500 scientists lift the lid on reproducibility. *Nature* 2016;**533**: 452-4.

Bakker BM, Bro C, Kotter P, Luttik MA, van Dijken JP, Pronk JT. The mitochondrial alcohol dehydrogenase Adh3p is involved in a redox shuttle in *Saccharomyces cerevisiae*. *J Bacteriol* 2000;**182**: 4730-7.

Bakker BM, Overkamp KM, van Maris AJ, Kotter P, Luttik MA, van Dijken JP, Pronk JT. Stoichiometry and compartmentation of NADH metabolism in *Saccharomyces cerevisiae*. *FEMS Microbiol Rev* 2001;**25**: 15-37.

Bao Z, Xiao H, Liang J, Zhang L, Xiong X, Sun N, . . . Zhao H. Homology-integrated CRISPR-Cas (HI-CRISPR) system for one-step multigene disruption in *Saccharomyces cerevisiae*. *ACS Synth Biol* 2015;**4**: 585-94.

Barlow JH, Rothstein R. Timing is everything: cell cycle control of Rad52. *Cell Div* 2010;**5**: 7.

Barnard H, Dooley AN, Areshian G, Gasparyan B, Faull KF. Chemical evidence for wine production around 4000 BCE in the Late Chalcolithic Near Eastern highlands. *Journal of Archaeological Science* 2011;**38**: 977-84.

Barnett JA. A history of research on yeasts. 1: Work by chemists and biologists 1789-1850. *Yeast* 1998;**14**: 1439-51.

Barnett JA. A history of research on yeasts 2: Louis Pasteur and his contemporaries, 1850-1880. *Yeast* 2000;**16**: 755-71.

Baudin A, Ozier-Kalogeropoulos O, Denouel A, Lacroute F, Cullin C. A simple and efficient method for direct gene deletion in *Saccharomyces cerevisiae*. *Nucleic Acids Res* 1993;**21**: 3329-30.

Blondin B, Gonde P, Ratomahenina R, Arnaud A, Galzy P. A study of cyanide-insensitive respiration in the genus *Dekkera* and *Brettanomyces*. *Microbiol Immunol* 1984;**28**: 637-44.

Boender LG, Almering MJ, Dijk M, van Maris AJ, de Winde JH, Pronk JT, Daran-Lapujade P. Extreme calorie restriction and energy source starvation in *Saccharomyces cerevisiae* represent distinct physiological states. *Biochim Biophys Acta* 2011;**1813**: 2133-44.

Boender LG, de Hulster EA, van Maris AJ, Daran-Lapujade PA, Pronk JT. Quantitative physiology of *Saccharomyces cerevisiae* at near-zero specific growth rates. *Appl Environ Microbiol* 2009;**75**: 5607-14.

Bonnet JA, de Kok HE, Roels JA. The growth of Saccharomyces cerevisiae CBS 426 on mixtures of glucose and ethanol: a model. *Antonie van Leeuwenhoek* 1980;**46**: 565-76.

Boulton SJ, Jackson SP. Identification of a *Saccharomyces cerevisiae* Ku80 homologue: roles in DNA double strand break rejoining and in telomeric maintenance. *Nucleic Acids Res* 1996;**24**: 4639-48.

Brickwedde A, Brouwers N, van den Broek M, Gallego Murillo JS, Fraiture JL, Pronk JT, Daran JG. Structural, Physiological and Regulatory Analysis of Maltose Transporter Genes in *Saccharomyces eubayanus* CBS 12357(T). *Front Microbiol* 2018;**9**: 1786.

Bridges HR, Fearnley IM, Hirst J. The subunit composition of mitochondrial NADH:ubiquinone oxidoreductase (complex I) from *Pichia pastoris*. *Mol Cell Proteomics* 2010;**9**: 2318-26.

Bridges HR, Grgic L, Harbour ME, Hirst J. The respiratory complexes I from the mitochondria of two *Pichia* species. *Biochem J* 2009;**422**: 151-9.

Brito N, Avila J, Perez MD, Gonzalez C, Siverio JM. The genes *YNI1* and *YNR1*, encoding nitrite reductase and nitrate reductase respectively in the yeast *Hansenula polymorpha*, are clustered and co-ordinately regulated. *Biochem J* 1996;**317 (Pt 1)**: 89-95.

Bruinenberg PM, van Dijken JP, Kuenen JG, Scheffers WA. Critical Parameters in the Isolation of Mitochondria from *Candida Utilis* Grown in Continuous Culture. *J Gen Microbiol* 1985a;**131**: 1035-42.

Bruinenberg PM, van Dijken JP, Kuenen JG, Scheffers WA. Oxidation of NADH and NADPH by mitochondria from the yeast *Candida utilis*. *J Gen Microbiol* 1985b;**131**: 1043-51.

Bruinenberg PM, van Dijken JP, Scheffers WA. An enzymic analysis of NADPH production and consumption in *Candida utilis*. *J Gen Microbiol* 1983a;**129**: 965-71.

Bruinenberg PM, van Dijken JP, Scheffers WA. A Theoretical Analysis of NADPH Production and Consumption in Yeasts. *J Gen Microbiol* 1983b;**129**: 953-64.

Buchholz K, Collins J. The roots--a short history of industrial microbiology and biotechnology. *Appl Microbiol Biotechnol* 2013;**97**: 3747-62.

Bud R. *Penicillin—triumph and tragedy.* Oxford: Oxford University Press, 2007.

Cabrera-Orefice A, Guerrero-Castillo S, Diaz-Ruiz R, Uribe-Carvajal S. Oxidative phosphorylation in *Debaryomyces hansenii*: physiological uncoupling at different growth phases. *Biochimie* 2014;**102**: 124-36.

Cagniard-Latour C. Mémoire sur la fermentation vineuse. *Ann Chim Phys* 1838;**68**: 206.

Cai P, Gao J, Zhou Y. CRISPR-mediated genome editing in non-conventional yeasts for biotechnological applications. *Microb Cell Fact* 2019;**18**: 63.

Canelas AB, Harrison N, Fazio A, Zhang J, Pitkanen JP, van den Brink J, . . . Nielsen J. Integrated multilaboratory systems biology reveals differences in protein metabolism between two reference yeast strains. *Nat Commun* 2010;**1**: 145.

Cao M, Gao M, Lopez-Garcia CL, Wu Y, Seetharam AS, Severin AJ, Shao Z. Centromeric DNA Facilitates Nonconventional Yeast Genetic Engineering. *ACS Synth Biol* 2017;**6**: 1545-53.

Carneiro P, Duarte M, Videira A. The main external alternative NAD(P)H dehydrogenase of *Neurospora crassa* mitochondria. *Biochim Biophys Acta* 2004;**1608**: 45-52.

Carneiro P, Duarte M, Videira A. The external alternative NAD(P)H dehydrogenase NDE3 is localized both in the mitochondria and in the cytoplasm of *Neurospora crassa*. *J Mol Biol* 2007;**368**: 1114-21.

Carus M, Dammer L. Food or Non-Food: Which Agricultural Feedstocks Are Best for Industrial Uses? *Industrial Biotechnology* 2013;**9**: 171-6.

Cereghino JL, Cregg JM. Heterologous protein expression in the methylotrophic yeast *Pichia pastoris*. *FEMS Microbiol Rev* 2000;**24**: 45-66.

Chance B, Williams GR. The respiratory chain and oxidative phosphorylation. *Adv Enzymol Relat Subj Biochem* 1956;**17**: 65-134.

Chapman JR, Taylor MR, Boulton SJ. Playing the end game: DNA double-strand break repair pathway choice. *Mol Cell* 2012;**47**: 497-510.

Chen Y, Nielsen J. Energy metabolism controls phenotypes by protein efficiency and allocation. *Proc Natl Acad Sci U S A* 2019;**116**: 17592-7.

Chesbro W, Evans T, Eifert R. Very slow growth of *Escherichia coli*. *J Bacteriol* 1979;**139**: 625-38.

Chung K-R, Lee M-H. Split-Marker-Mediated Transformation and Targeted Gene Disruption in Filamentous Fungi. In: van den Berg MA, Maruthachalam K (eds.) *Genetic Transformation Systems in Fungi, Volume 2*. Cham: Springer International Publishing, 2015, 175-80.

Cohen SN, Chang AC, Boyer HW, Helling RB. Construction of biologically functional bacterial plasmids in vitro. *Proc Natl Acad Sci U S A* 1973;**70**: 3240-4.

Conesa A, Nueda MJ, Ferrer A, Talon M. maSigPro: a method to identify significantly differential expression profiles in time-course microarray experiments. *Bioinformatics* 2006;**22**: 1096-102.

Cong L, Ran FA, Cox D, Lin S, Barretto R, Habib N, . . . Zhang F. Multiplex genome engineering using CRISPR/Cas systems. *Science* 2013;**339**: 819-23.

Correia K, Yu SM, Mahadevan R. Reconstructing the evolution of metabolism in budding yeasts. *bioRxiv* 2017: 237974.

Dashko S, Zhou N, Compagno C, Piskur J. Why, when, and how did yeast evolve alcoholic fermentation? *FEMS Yeast Res* 2014;**14**: 826-32.

De Deken RH. The Crabtree effect: a regulatory system in yeast. *J Gen Microbiol* 1966;**44**: 149-56.

de Koning W, Harder W, Dijkhuizen L. Glycerol metabolism in the methylotrophic yeast *Hansenula polymorpha*: phosphorylation as the initial step. *Arch Microbiol* 1987;**148**: 314-20.

de Vries S, Marres CA. The mitochondrial respiratory chain of yeast. Structure and biosynthesis and the role in cellular metabolism. *Biochim Biophys Acta* 1987;**895**: 205-39.

Deaner M, Mejia J, Alper HS. Enabling Graded and Large-Scale Multiplex of Desired Genes Using a Dual-Mode dCas9 Activator in *Saccharomyces cerevisiae*. *ACS Synth Biol* 2017;**6**: 1931-43.

Delneri D, Tomlin GC, Wixon JL, Hutter A, Sefton M, Louis EJ, Oliver SG. Exploring redundancy in the yeast genome: an improved strategy for use of the cre-loxP system. *Gene* 2000;**252**: 127-35.

Desplats C, Beyly A, Cuine S, Bernard L, Cournac L, Peltier G. Modification of substrate specificity in single point mutants of *Agrobacterium tumefaciens* type II NADH dehydrogenase. *FEBS Lett* 2007;**581**: 4017-22.

Di Primio C, Galli A, Cervelli T, Zoppe M, Rainaldi G. Potentiation of gene targeting in human cells by expression of *Saccharomyces cerevisiae* Rad52. *Nucleic Acids Res* 2005;**33**: 4639-48.

DiCarlo JE, Norville JE, Mali P, Rios X, Aach J, Church GM. Genome engineering in *Saccharomyces cerevisiae* using CRISPR-Cas systems. *Nucleic Acids Res* 2013;**41**: 4336-43.

Diderich JA, Schepper M, van Hoek P, Luttik MA, van Dijken JP, Pronk JT, . . . Kruckeberg AL. Glucose uptake kinetics and transcription of HXT genes in chemostat cultures of *Saccharomyces cerevisiae*. *J Biol Chem* 1999;**274**: 15350-9.

Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, . . . Gingeras TR. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* 2013;**29**: 15-21.

Dominiak K, Koziel A, Jarmuszkiewicz W. The interplay between mitochondrial reactive oxygen species formation and the coenzyme Q reduction level. *Redox Biol* 2018;**18**: 256-65.

Doran PM. Chapter 10 - Mass Transfer. In: Doran PM (ed.) *Bioprocess Engineering Principles (Second Edition)*. London: Academic Press, 2013, 379-444.

Duarte M, Peters M, Schulte U, Videira A. The internal alternative NADH dehydrogenase of *Neurospora crassa* mitochondria. *Biochem J* 2003;**371**: 1005-11.

Dudley R. Ethanol, fruit ripening, and the historical origins of human alcoholism in primate frugivory. *Integr Comp Biol* 2004;**44**: 315-23.

Dujon B. Yeast evolutionary genomics. *Nat Rev Genet* 2010;**11**: 512-24.

Edgar RC. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res* 2004;**32**: 1792-7.

Egli T, Quayle JR. Influence of the Carbon-Nitrogen Ratio of the Growth Medium on the Cellular Composition and the Ability of the Methylotrophic Yeast *Hansenula polymorpha* to Utilize Mixed Carbon-Sources. *J Gen Microbiol* 1986;**132**: 1779-88.

Egli T, Van Dijken JP, Veenhuis M, Harder W, Fiechter A. Methanol Metabolism in Yeasts - Regulation of the Synthesis of Catabolic Enzymes. *Arch Microbiol* 1980;**124**: 115-21.

Eldarov MA, Mardanov AV, Beletsky AV, Ravin NV, Skryabin KG. Complete sequence and analysis of the mitochondrial genome of the methylotrophic yeast *Hansenula polymorpha* DL-1. *FEMS Yeast Res* 2011;**11**: 464-72.

Engler C, Kandzia R, Marillonnet S. A one pot, one step, precision cloning method with high throughput capability. *PloS one* 2008;**3**: e3647.

Ercan O, Bisschops MM, Overkamp W, Jorgensen TR, Ram AF, Smid EJ, . . . Kleerebezem M. Physiological and Transcriptional Responses of Different Industrial Microbes at Near-Zero Specific Growth Rates. *Appl Environ Microbiol* 2015;**81**: 5662-70.

Ereky K. *Biotechnologie der Fleisch-, Fett-, und Milcherzeugung im landwirtschaftlichen Grossbetriebe: für naturwissenschaftlich gebildete Landwirte verfasst*. Berlin: P. Parey, 1919.

Es I, Mousavi Khaneghah A, Barba FJ, Saraiva JA, Sant'Ana AS, Hashemi SMB. Recent advancements in lactic acid production - a review. *Food Res Int* 2018;**107**: 763-70.

Eubel H, Heinemeyer J, Sunderhaus S, Braun HP. Respiratory chain supercomplexes in plant mitochondria. *Plant physiology and biochemistry : PPB* 2004;**42**: 937-42.

Fairhead C, Llorente B, Denis F, Soler M, Dujon B. New vectors for combinatorial deletions in yeast chromosomes and for gap-repair cloning using 'split-marker' recombination. *Yeast* 1996;**12**: 1439-57.

Fecke W, Sled VD, Ohnishi T, Weiss H. Disruption of the gene encoding the NADH-binding subunit of NADH: ubiquinone oxidoreductase in *Neurospora crassa*. Formation of a partially assembled enzyme without FMN and the iron-sulphur cluster N-3. *Eur J Biochem* 1994;**220**: 551-8.

Feng Y, Li W, Li J, Wang J, Ge J, Xu D, . . . Yang M. Structural insight into the type-II mitochondrial NADH dehydrogenases. *Nature* 2012;**491**: 478-82.

Ferguson SJ. ATP synthase: from sequence to ring size to the P/O ratio. *Proc Natl Acad Sci U S A* 2010;**107**: 16755-6.

Ferrer-Miralles N, Domingo-Espin J, Corchero JL, Vazquez E, Villaverde A. Microbial factories for recombinant pharmaceuticals. *Microb Cell Fact* 2009;**8**: 17.

Forti L, Di Mauro S, Cramarossa MR, Filippucci S, Turchetti B, Buzzini P. Non-Conventional Yeasts Whole Cells as Efficient Biocatalysts for the Production of Flavors and Fragrances. *Molecules* 2015;**20**: 10377-98.

Fraczek MG, Naseeb S, Delneri D. History of genome editing in yeast. *Yeast* 2018;**35**: 361-8.

Freedman LP, Cockburn IM, Simcoe TS. The economics of reproducibility in preclinical research. *PLoS Biol* 2015;**13**: e1002165.

Fromm S, Senkler J, Eubel H, Peterhansel C, Braun HP. Life without complex I: proteome analyses of an *Arabidopsis* mutant lacking the mitochondrial NADH dehydrogenase complex. *J Exp Bot* 2016;**67**: 3079-93.

Fukasawa Y, Tsuji J, Fu SC, Tomii K, Horton P, Imai K. MitoFates: improved prediction of mitochondrial targeting sequences and their cleavage sites. *Mol Cell Proteomics* 2015;**14**: 1113-26.

Gao S, Tong Y, Wen Z, Zhu L, Ge M, Chen D, . . . Yang S. Multiplex gene editing of the *Yarrowia lipolytica* genome using the CRISPR-Cas9 system. *J Ind Microbiol Biotechnol* 2016;**43**: 1085-93.

Gao Y, Zhao Y. Self-processing of ribozyme-flanked RNAs into guide RNAs in vitro and in vivo for CRISPR-mediated genome editing. *J Integr Plant Biol* 2014;**56**: 343-9.

Gardner TS. A swan in the making. *Science* 2014;**345**: 855.

Gardner TS. Systems and methods for process design and analysis. *International patent* WO/2016/019188 A1. 2016.

Gasser SM, Daum G, Schatz G. Import of proteins into mitochondria. Energy-dependent uptake of precursors by isolated mitochondria. *J Biol Chem* 1982;**257**: 13034-41.

Gellissen G. Heterologous protein production in methylotrophic yeasts. *Appl Microbiol Biotechnol* 2000;**54**: 741-50.

Gibson DG, Young L, Chuang RY, Venter JC, Hutchison CA, 3rd, Smith HO. Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nat Methods* 2009;**6**: 343-5.

Gietz RD, Schiestl RH. High-efficiency yeast transformation using the LiAc/SS carrier DNA/PEG method. *Nat Protoc* 2007;**2**: 31-4.

Gilkerson RW, Selker JM, Capaldi RA. The cristal membrane of mitochondria is the principal site of oxidative phosphorylation. *FEBS Lett* 2003;**546**: 355-8.

Godoy-Hernandez A, Tate DJ, McMillan DGG. Revealing the Membrane-Bound Catalytic Oxidation of NADH by the Drug Target Type-II NADH Dehydrogenase. *Biochemistry* 2019;**58**: 4272-5.

Goffeau A, Barrell BG, Bussey H, Davis RW, Dujon B, Feldmann H, . . . Oliver SG. Life with 6000 genes. *Science* 1996;**274**: 546, 63-7.

Goldstein AL, McCusker JH. Three new dominant drug resistance cassettes for gene disruption in *Saccharomyces cerevisiae*. *Yeast* 1999;**15**: 1541-53.

Gonzalez-Barroso MM, Ledesma A, Lepper S, Perez-Magan E, Zaragoza P, Rial E. Isolation and bioenergetic characterization of mitochondria from *Pichia pastoris*. *Yeast* 2006;**23**: 307-13.

Goodwin S, McPherson JD, McCombie WR. Coming of age: ten years of next-generation sequencing technologies. *Nat Rev Genet* 2016;**17**: 333-51.

Gorter de Vries AR, de Groot PA, van den Broek M, Daran JG. CRISPR-Cas9 mediated gene deletions in lager yeast *Saccharomyces pastorianus*. *Microb Cell Fact* 2017;**16**: 222.

Guerrero-Castillo S, Araiza-Olivera D, Cabrera-Orefice A, Espinasa-Jaramillo J, Gutierrez-Aguilar M, Luevano-Martinez LA, . . . Uribe-Carvajal S. Physiological uncoupling of mitochondrial oxidative phosphorylation. Studies in different yeast species. *J Bioenerg Biomembr* 2011;**43**: 323-31.

Guerrero-Castillo S, Baertling F, Kownatzki D, Wessels HJ, Arnold S, Brandt U, Nijtmans L. The Assembly Pathway of Mitochondrial Respiratory Chain Complex I. *Cell Metab* 2017;**25**: 128-39.

Guerrero-Castillo S, Cabrera-Orefice A, Vazquez-Acevedo M, Gonzalez-Halphen D, Uribe-Carvajal S. During the stationary growth phase, *Yarrowia lipolytica* prevents the overproduction of reactive oxygen species by activating an uncoupled mitochondrial respiratory pathway. *Biochim Biophys Acta* 2012;**1817**: 353-62.

Hagman A, Sall T, Compagno C, Piskur J. Yeast "make-accumulate-consume" life strategy evolved as a multi-step process that predates the whole genome duplication. *PloS one* 2013;**8**: e68734.

Hakkaart XDV, Pronk JT, van Maris AJA. A simulator-assisted workshop for teaching chemostat cultivation in academic classes on microbial physiology. *J Microbiol Biol Educ* 2017;**18**.

Hensing MC, Rouwenhorst RJ, Heijnen JJ, van Dijken JP, Pronk JT. Physiological and technological aspects of large-scale heterologous-protein production with yeasts. *Antonie van Leeuwenhoek* 1995;**67**: 261-79.

Hinkle PC. P/O ratios of mitochondrial oxidative phosphorylation. *Biochim Biophys Acta* 2005;**1706**: 1-11.

Hinkle PC, Kumar MA, Resetar A, Harris DL. Mechanistic stoichiometry of mitochondrial oxidative phosphorylation. *Biochemistry* 1991;**30**: 3576-82.

Hinnen A, Hicks JB, Fink GR. Transformation of yeast. *Proc Natl Acad Sci U S A* 1978;**75**: 1929-33.

Hirst J. Open questions: respiratory chain supercomplexes-why are they there and what do they do? *BMC Biol* 2018;**16**: 111.

Hong KK, Hou J, Shoaie S, Nielsen J, Bordel S. Dynamic 13C-labeling experiments prove important differences in protein turnover rate between two *Saccharomyces cerevisiae* strains. *FEMS Yeast Res* 2012;**12**: 741-7.

Horwitz AA, Walter JM, Schubert MG, Kung SH, Hawkins K, Platt DM, . . . Newman JD. Efficient Multiplexed Integration of Synergistic Alleles and Metabolic Pathways in Yeasts via CRISPR-Cas. *Cell Syst* 2015;**1**: 88-96.

Hsu PD, Lander ES, Zhang F. Development and applications of CRISPR-Cas9 for genome engineering. *Cell* 2014;**157**: 1262-78.

Hunte C, Zickermann V, Brandt U. Functional modules and structural basis of conformational coupling in mitochondrial complex I. *Science* 2010;**329**: 448-51.

Intern.YeastCo.Ltd. *Verfahren zur Herstellung von Hefe nach dem Zulaufverfahren* (Germany Patent No. DE583760C), 1933.

Ioannidis JP. Why most published research findings are false. *PLoS Med* 2005;**2**: e124.

Iwata M, Lee Y, Yamashita T, Yagi T, Iwata S, Cameron AD, Maher MJ. The structure of the yeast NADH dehydrogenase (Ndi1) reveals overlapping binding sites for water- and lipid-soluble substrates. *Proc Natl Acad Sci U S A* 2012;**109**: 15247-52.

Jansen MLA, Bracher JM, Papapetridis I, Verhoeven MD, de Bruijn H, de Waal PP, . . . Pronk JT. Saccharomyces cerevisiae strains for second-generation ethanol production: from academic exploration to industrial implementation. *FEMS Yeast Res* 2017;**17**.

Jasin M, Rothstein R. Repair of strand breaks by homologous recombination. *Cold Spring Harb Perspect Biol* 2013;**5**: a012740.

Jinek M, Chylinski K, Fonfara I, Hauer M, Doudna JA, Charpentier E. A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science* 2012;**337**: 816-21.

Johnson EA. Biotechnology of non-*Saccharomyces* yeasts-the ascomycetes. *Appl Microbiol Biotechnol* 2013;**97**: 503-17.

Jones AJ, Blaza JN, Varghese F, Hirst J. Respiratory Complex I in *Bos taurus* and *Paracoccus denitrificans* Pumps Four Protons across the Membrane for Every NADH Oxidized. *J Biol Chem* 2017;**292**: 4987-95.

Joseph-Horne T, Hollomon DW, Wood PM. Fungal respiration: a fusion of standard and alternative components. *Biochim Biophys Acta* 2001;**1504**: 179-95.

Juergens H, Hakkaart XDV, Bras JE, Vente A, Wu L, Benjamin KR, . . . Mans R. Contribution of Complex I NADH Dehydrogenase to Respiratory Energy Coupling in Glucose-Grown Cultures of *Ogataea parapolymorpha*. *Appl Environ Microbiol* 2020a;**86**.

Juergens H, Hakkaart XDV, Bras JE, Vente A, Wu L, Benjamin KR, . . . Mans R. Proteomics data of glucose-grown *Ogataea parapolymorpha*, DOI 10.6084/m9.figshare.11398773. Figshare, 2020b.

Juergens H, Niemeijer M, Jennings-Antipov LD, Mans R, Morel J, van Maris AJA, . . . Gardner TS. Evaluation of a novel cloud-based software platform for structured experiment design and linked data analytics. *Sci Data* 2018a;**5**: 180195.

Juergens H, Varela JA, Gorter de Vries AR, Perli T, Gast VJM, Gyurchev NY, . . . Daran JG. Genome editing in *Kluyveromyces* and *Ogataea* yeasts using a broad-host-range Cas9/gRNA co-expression plasmid. *FEMS Yeast Res* 2018b;**18**.

Kalvala A, Rainaldi G, Di Primio C, Liverani V, Falaschi A, Galli A. Enhancement of gene targeting in human cells by intranuclear permeation of the *Saccharomyces cerevisiae* Rad52 protein. *Nucleic Acids Res* 2010;**38**: e149.

Kang HA, Kang W, Hong WK, Kim MW, Kim JY, Sohn JH, . . . Rhee SK. Development of expression systems for the production of recombinant human serum albumin using the *MOX* promoter in *Hansenula polymorpha* DL-1. *Biotechnol Bioeng* 2001;**76**: 175-85.

Karanam K, Kafri R, Loewer A, Lahav G. Quantitative live cell imaging reveals a gradual shift between DNA repair mechanisms and a maximal use of HR in mid S phase. *Mol Cell* 2012;**47**: 320-9.

Karp H, Alamae T. Glucose transport in a methylotrophic yeast *Hansenula polymorpha*. *FEMS Microbiol Lett* 1998;**166**: 267-73.

Kata I, Semkiv MV, Ruchala J, Dmytruk KV, Sibirny AA. Overexpression of the genes *PDC1* and *ADH1* activates glycerol conversion to ethanol in the thermotolerant yeast *Ogataea* (*Hansenula*) *polymorpha*. *Yeast* 2016;**33**: 471-8.

Katz R, Kilpatrick L, Chance B. Acquisition and loss of rotenone sensitivity in *Torulopsis utilis*. *Eur J Biochem* 1971;**21**: 301-7.

Kavscek M, Strazar M, Curk T, Natter K, Petrovic U. Yeast as a cell factory: current state and perspectives. *Microb Cell Fact* 2015;**14**: 94.

Kerscher S, Drose S, Zickermann V, Brandt U. The three families of respiratory NADH dehydrogenases. *Results Probl Cell Differ* 2008;**45**: 185-222.

Kerscher SJ. Diversity and origin of alternative NADH:ubiquinone oxidoreductases. *Biochim Biophys Acta* 2000;**1459**: 274-83.

Kerscher SJ, Okun JG, Brandt U. A single external enzyme confers alternative NADH:ubiquinone oxidoreductase activity in *Yarrowia lipolytica*. *J Cell Sci* 1999;**112 (Pt 14)**: 2347-54.

Kiers J, Zeeman AM, Luttik M, Thiele C, Castrillo JI, Steensma HY, . . . Pronk JT. Regulation of alcoholic fermentation in batch and chemostat cultures of *Kluyveromyces lactis* CBS 2359. *Yeast* 1998;**14**: 459-69.

Kim DU, Hayles J, Kim D, Wood V, Park HO, Won M, . . . Hoe KL. Analysis of a genome-wide set of gene deletions in the fission yeast *Schizosaccharomyces pombe*. *Nat Biotechnol* 2010;**28**: 617-23.

Kitajima-Ihara T, Yagi T. Rotenone-insensitive internal NADH-quinone oxidoreductase of *Saccharomyces cerevisiae* mitochondria: the enzyme expressed in *Escherichia coli* acts as a member of the respiratory chain in the host cells. *FEBS Lett* 1998;**421**: 37-40.

Klinner U, Schäfer B. Genetic aspects of targeted insertion mutagenesis in yeasts. *FEMS Microbiol Rev* 2004;**28**: 201-23.

Knapp F. *Lehrbuch der chemischen Technologie* volume 2. Braunschweig: F. Vieweg und Sohn, 1847.

Knijnenburg TA, Daran JM, van den Broek MA, Daran-Lapujade PA, de Winde JH, Pronk JT, . . . Wessels LF. Combinatorial effects of environmental parameters on transcriptional regulation in *Saccharomyces cerevisiae*: a quantitative analysis of a compendium of chemostat-based transcriptome data. *BMC Genomics* 2009;**10**: 53.

Kooistra R, Hooykaas PJ, Steensma HY. Efficient gene targeting in *Kluyveromyces lactis*. *Yeast* 2004;**21**: 781-92.

Kuijpers NG, Chroumpi S, Vos T, Solis-Escalante D, Bosman L, Pronk JT, . . . Daran-Lapujade P. One-step assembly and targeted integration of multigene constructs assisted by the I-SceI meganuclease in *Saccharomyces cerevisiae*. *FEMS Yeast Res* 2013a;**13**: 769-81.

Kuijpers NG, Solis-Escalante D, Bosman L, van den Broek M, Pronk JT, Daran JM, Daran-Lapujade P. A versatile, efficient strategy for assembly of multi-fragment expression vectors in *Saccharomyces cerevisiae* using 60 bp synthetic recombination sequences. *Microb Cell Fact* 2013b;**12**: 47.

Kunze G, Kang HA, Gellissen G. *Hansenula polymorpha* (*Pichia angusta*): biology and applications. In: Satyanarayana T, Kunze G (eds.) *Yeast Biotechnology: Diversity and Applications* volume 1. Dordrecht: Springer Netherlands, 2009.

Kurtzman CP. *Ogataea* Y. Yamada, K. Maeda & Mikata (1994). In: Kurtzman CP, Fell JW (eds.) *The Yeasts (Fifth Edition)* volume 1. London: Elsevier, 2011.

Kurylenko OO, Ruchala J, Hryniv OB, Abbas CA, Dmytruk KV, Sibirny AA. Metabolic engineering and classical selection of the methylotrophic thermotolerant yeast *Hansenula polymorpha* for improvement of high-temperature xylose alcoholic fermentation. *Microb Cell Fact* 2014;**13**: 122.

Kützing F. Microscopische Untersuchungen über die Hefe und Essigmutter, nebst mehreren andern dazu gehörigen vegetabilischen Gebilden. *J Prakt Chem* 1837;**11**: 385-409.

Laroche T, Martin SG, Gotta M, Gorham HC, Pryde FE, Louis EJ, Gasser SM. Mutation of yeast Ku genes disrupts the subnuclear organization of telomeres. *Curr Biol* 1998;**8**: 653-6.

Larsson C, Pahlman IL, Ansell R, Rigoulet M, Adler L, Gustafsson L. The importance of the glycerol 3-phosphate shuttle during aerobic growth of *Saccharomyces cerevisiae*. *Yeast* 1998;**14**: 347-57.

Lee ME, DeLoache WC, Cervantes B, Dueber JE. A Highly Characterized Yeast Toolkit for Modular, Multipart Assembly. *ACS Synth Biol* 2015;**4**: 975-86.

Levine DW, Cooney CL. Isolation and characterization of a thermotolerant methanol-utilizing yeast. *Appl Microbiol* 1973;**26**: 982-90.

Li H, Durbin R. Fast and accurate long-read alignment with Burrows-Wheeler transform. *Bioinformatics* 2010;**26**: 589-95.

Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, . . . Genome Project Data Processing S. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* 2009;**25**: 2078-9.

Li M, Borodina I. Application of synthetic biology for production of chemicals in yeast *Saccharomyces cerevisiae*. *FEMS Yeast Res* 2015;**15**: 1-12.

Liachko I, Dunham MJ. An autonomously replicating sequence for use in a wide range of budding yeasts. *FEMS Yeast Res* 2014;**14**: 364-7.

Liu L, Wang J, Rosenberg D, Zhao H, Lengyel G, Nadel D. Fermented beverage and food storage in 13,000 y-old stone mortars at Raqefet Cave, Israel: Investigating Natufian ritual feasting. *J Archaeol Sci Rep* 2018;**21**: 783-93.

Löbs A-K, Engel R, Schwartz C, Flores A, Wheeldon I. CRISPR-Cas9-enabled genetic disruptions for understanding ethanol and ethyl acetate biosynthesis in *Kluyveromyces marxianus*. *Biotechnol Biofuels* 2017a;**10**: 164.

Löbs AK, Schwartz C, Wheeldon I. Genome and metabolic engineering in non-conventional yeasts: Current advances and applications. *Synth Syst Biotechnol* 2017b;**2**: 198-207.

Lõoke M, Kristjuhan K, Kristjuhan A. Extraction of genomic DNA from yeasts for PCR-based applications. *Biotechniques* 2011;**50**: 325-8.

Lorenz R, Bernhart SH, Honer Zu Siederdissen C, Tafer H, Flamm C, Stadler PF, Hofacker IL. ViennaRNA Package 2.0. *Algorithms Mol Biol* 2011;**6**: 26.

Lowry OH, Rosebrough NJ, Farr AL, Randall RJ. Protein measurement with the Folin phenol reagent. *J Biol Chem* 1951;**193**: 265-75.

Luttik MA, Kotter P, Salomons FA, van der Klei IJ, van Dijken JP, Pronk JT. The *Saccharomyces cerevisiae ICL2* gene encodes a mitochondrial 2-methylisocitrate lyase involved in propionyl-coenzyme A metabolism. *J Bacteriol* 2000;**182**: 7007-13.

Luttik MA, Overkamp KM, Kotter P, de Vries S, van Dijken JP, Pronk JT. The *Saccharomyces cerevisiae NDE1* and *NDE2* genes encode separate mitochondrial NADH dehydrogenases catalyzing the oxidation of cytosolic NADH. *J Biol Chem* 1998;**273**: 24529-34.

Madigan MT, Martinko JM, Bender KS, Buckley DH, Stahl DA. *Brock biology of microorganisms*, 2014.

Mali P, Yang L, Esvelt KM, Aach J, Guell M, DiCarlo JE, . . . Church GM. RNA-guided human genome engineering via Cas9. *Science* 2013;**339**: 823-6.

Manfrao-Netto JHC, Gomes AMV, Parachin NS. Advances in Using *Hansenula polymorpha* as Chassis for Recombinant Protein Production. *Front Bioeng Biotechnol* 2019;**7**: 94.

Mans R, van Rossum HM, Wijsman M, Backx A, Kuijpers NG, van den Broek M, . . . Daran JM. CRISPR/Cas9: a molecular Swiss army knife for simultaneous introduction of multiple genetic modifications in *Saccharomyces cerevisiae*. *FEMS Yeast Res* 2015;**15**: fov004.

Marques I, Dencher NA, Videira A, Krause F. Supramolecular organization of the respiratory chain in *Neurospora crassa* mitochondria. *Eukaryot Cell* 2007;**6**: 2391-405.

Marreiros BC, Calisto F, Castro PJ, Duarte AM, Sena FV, Silva AF, . . . Pereira MM. Exploring membrane respiratory chains. *Biochim Biophys Acta* 2016a;**1857**: 1039-67.

Marreiros BC, Sena FV, Sousa FM, Batista AP, Pereira MM. Type II NADH:quinone oxidoreductase family: phylogenetic distribution, structural diversity and evolutionary divergences. *Environ Microbiol* 2016b;**18**: 4697-709.

Mashego MR, van Gulik WM, Vinke JL, Heijnen JJ. Critical evaluation of sampling techniques for residual glucose determination in carbon-limited chemostat culture of *Saccharomyces cerevisiae*. *Biotechnol Bioeng* 2003;**83**: 395-9.

Mattanovich D, Branduardi P, Dato L, Gasser B, Sauer M, Porro D. Recombinant protein production in yeasts. *Methods Mol Biol* 2012;**824**: 329-58.

Mattanovich D, Sauer M, Gasser B. Yeast biotechnology: teaching the old dog new tricks. *Microb Cell Fact* 2014;**13**: 34.

Matus-Ortega MG, Cardenas-Monroy CA, Flores-Herrera O, Mendoza-Hernandez G, Miranda M, Gonzalez-Pedrajo B, . . . Pardo JP. New complexes containing the internal alternative NADH dehydrogenase (Ndi1) in mitochondria of *Saccharomyces cerevisiae*. *Yeast* 2015;**32**: 629-41.

Mazzoni C, Serafini A, Falcone C. The inactivation of *KlNOT4*, a *Kluyveromyces lactis* gene encoding a component of the CCR4-NOT complex, reveals new regulatory functions. *Genetics* 2005;**170**: 1023-32.

Meadows AL, Hawkins KM, Tsegaye Y, Antipov E, Kim Y, Raetz L, . . . Tsong AE. Rewriting yeast central carbon metabolism for industrial isoprenoid production. *Nature* 2016;**537**: 694-7.

Melo AM, Bandeiras TM, Teixeira M. New insights into type II NAD(P)H:quinone oxidoreductases. *Microbiol Mol Biol Rev* 2004;**68**: 603-16.

Melo AM, Duarte M, Moller IM, Prokisch H, Dolan PL, Pinto L, . . . Videira A. The external calcium-dependent NADPH dehydrogenase from *Neurospora crassa* mitochondria. *J Biol Chem* 2001;**276**: 3947-51.

Meyen FJF. Jahresbericht über die Resultate der Arbeiten im Felde der physiologischen Botanik von dem Jahre 1837. *Arch Naturgeschich* 1838;**4**: 1-186.

Meyer H-P, Minas W, Schmidhalter D. Industrial-Scale Fermentation. In: Liao JC, Wittmann C (eds.) *Industrial Biotechnology: Products and Processes*: Wiley-VCH Verlag GmbH & Co. KGaA, 2016, 1-53.

Michalecka AM, Agius SC, Moller IM, Rasmusson AG. Identification of a mitochondrial external NADPH dehydrogenase by overexpression in transgenic *Nicotiana sylvestris*. *Plant J* 2004;**37**: 415-25.

Mitchell P. Chemiosmotic coupling in oxidative and photosynthetic phosphorylation. *Biol Rev Camb Philos Soc* 1966;**41**: 445-502.

Molenaar D, van Berlo R, de Ridder D, Teusink B. Shifts in growth strategies reflect tradeoffs in cellular economics. *Mol Syst Biol* 2009;**5**: 323.

Morrissey JP, Etschmann MM, Schrader J, de Billerbeck GM. Cell factory applications of the yeast *Kluyveromyces marxianus* for the biotechnological production of natural flavour and fragrance molecules. *Yeast* 2015;**32**: 3-16.

Mullis K, Faloona F, Scharf S, Saiki R, Horn G, Erlich H. Specific enzymatic amplification of DNA in vitro: the polymerase chain reaction. 1986. *Biotechnology* 1992;**24**: 17-27.

Nambu-Nishida Y, Nishida K, Hasunuma T, Kondo A. Development of a comprehensive set of tools for genome engineering in a cold- and thermo-tolerant *Kluyveromyces marxianus* yeast strain. *Sci Rep* 2017;**7**: 8993.

Nantapong N, Otofuji A, Migita CT, Adachi O, Toyama H, Matsushita K. Electron transfer ability from NADH to menaquinone and from NADPH to oxygen of type II NADH dehydrogenase of *Corynebacterium glutamicum*. *Biosci Biotechnol Biochem* 2005;**69**: 149-59.

Navarro FJ, Perdomo G, Tejera P, Medina B, Machin F, Guillen RM, . . . Siverio JM. The role of nitrate reductase in the regulation of the nitrate assimilation pathway in the yeast *Hansenula polymorpha*. *FEMS Yeast Res* 2003;**4**: 149-55.

Nes WD. Biosynthesis of cholesterol and other sterols. *Chem Rev* 2011;**111**: 6423-51.

Nielsen JB, Nielsen ML, Mortensen UH. Transient disruption of non-homologous end-joining facilitates targeted genome manipulations in the filamentous fungus *Aspergillus nidulans*. *Fungal Genet Biol* 2008;**45**: 165-70.

Nilsson A, Bjornson E, Flockhart M, Larsen FJ, Nielsen J. Complex I is bypassed during high intensity exercise. *Nat Commun* 2019;**10**: 5072.

Nonklang S, Abdel-Banat BM, Cha-aim K, Moonjai N, Hoshida H, Limtong S, . . . Akada R. High-temperature ethanol fermentation and transformation with linear DNA in the thermotolerant yeast *Kluyveromyces marxianus* DMKU3-1042. *Appl Environ Microbiol* 2008;**74**: 7514-21.

Noy NF, Shah NH, Whetzel PL, Dai B, Dorf M, Griffith N, . . . Musen MA. BioPortal: ontologies and integrated data resources at the click of a mouse. *Nucleic Acids Res* 2009;**37**: W170-3.

Nubel E, Wittig I, Kerscher S, Brandt U, Schagger H. Two-dimensional native electrophoretic analysis of respiratory supercomplexes from *Yarrowia lipolytica*. *Proteomics* 2009;**9**: 2408-18.

Nueda MJ, Tarazona S, Conesa A. Next maSigPro: updating maSigPro bioconductor package for RNA-seq time series. *Bioinformatics* 2014;**30**: 2598-602.

Numamoto M, Maekawa H, Kaneko Y. Efficient genome editing by CRISPR/Cas9 with a tRNA-sgRNA fusion in the methylotrophic yeast *Ogataea polymorpha*. *J Biosci Bioeng* 2017;**124**: 487-92.

NWO. Nederlandse Organisatie voor Wetenschappelijk Onderzoek. NWO makes 3 million available for Replication Studies pilot. 2016 [Available from: https://www.nwo.nl/en/news-and-events/news/2016/nwo-makes-3-million-available-for-replication-studies-pilot.html].

OECD. "Biotechnology", in *OECD Factbook 2013: Economic, Environmental and Social Statistics*, OECD Publishing, Paris. 2013.

Orr-Weaver TL, Szostak JW, Rothstein RJ. Yeast transformation: a model system for the study of recombination. *Proc Natl Acad Sci U S A* 1981;**78**: 6354-8.

Overkamp KM, Bakker BM, Kotter P, van Tuijl A, de Vries S, van Dijken JP, Pronk JT. In vivo analysis of the mechanisms for oxidation of cytosolic NADH by *Saccharomyces cerevisiae* mitochondria. *J Bacteriol* 2000;**182**: 2823-30.

Overkamp KM, Bakker BM, Steensma HY, van Dijken JP, Pronk JT. Two mechanisms for oxidation of cytosolic NADPH by *Kluyveromyces lactis* mitochondria. *Yeast* 2002;**19**: 813-24.

Palmieri F, Agrimi G, Blanco E, Castegna A, Di Noia MA, Iacobazzi V, . . . Walker J. Identification of mitochondrial carriers in *Saccharomyces cerevisiae* by transport assay of reconstituted recombinant proteins. *Biochim Biophys Acta* 2006;**1757**: 1249-62.

Pasteur L. Ueber die alkoholische Gährung. *J Prakt Chem* 1858a;**73**: 451-5.

Pasteur L. Ueber die Milchsäuregährung. *J Prakt Chem* 1858b;**73**: 447-51.

Pasteur L. *Improvement in the manufacture of beer and yeast* (US Patent No. 141072), 1873.

Patchett RA, Jones CW. The apparent oxidation of NADH by whole cells of the methylotrophic bacterium *Methylophilus methylotrophus*. A cautionary tale. *Antonie van Leeuwenhoek* 1986;**52**: 387-92.

Pfeiffer T, Schuster S, Bonhoeffer S. Cooperation and competition in the evolution of ATP-producing pathways. *Science* 2001;**292**: 504-7.

Pignocchi C, Berardi E, Cox BS. Nitrate reduction and the isolation of Nit$^-$ mutants in *Hansenula polymorpha*. *Microbiology* 1998;**144**: 2323-30.

Piper MD, Daran-Lapujade P, Bro C, Regenberg B, Knudsen S, Nielsen J, Pronk JT. Reproducibility of oligonucleotide microarray transcriptome analyses. An interlaboratory comparison using chemostat cultures of *Saccharomyces cerevisiae*. *J Biol Chem* 2002;**277**: 37001-8.

Pirt SJ. The maintenance energy of bacteria in growing cultures. *Proc R Soc Lond B Biol Sci* 1965;**163**: 224-31.

Plessis A, Dujon B. Multiple tandem integrations of transforming DNA sequences in yeast chromosomes suggest a mechanism for integrative transformation by homologous recombination. *Gene* 1993;**134**: 41-50.

Polotnianka RM, Li J, Lustig AJ. The yeast Ku heterodimer is essential for protection of the telomere against nucleolytic and recombinational activities. *Curr Biol* 1998;**8**: 831-4.

Porro D, Gasser B, Fossati T, Maurer M, Branduardi P, Sauer M, Mattanovich D. Production of recombinant proteins and metabolites in yeasts: when are these systems better than bacterial production systems? *Appl Microbiol Biotechnol* 2011;**89**: 939-48.

Postma E, Verduyn C, Scheffers WA, Van Dijken JP. Enzymic analysis of the Crabtree effect in glucose-limited chemostat cultures of *Saccharomyces cerevisiae*. *Appl Environ Microbiol* 1989;**55**: 468-77.

Prömper C, Schneider R, Weiss H. The role of the proton-pumping and alternative respiratory chain NADH:ubiquinone oxidoreductases in overflow catabolism of *Aspergillus niger*. *Eur J Biochem* 1993;**216**: 223-30.

Radecka D, Mukherjee V, Mateo RQ, Stojiljkovic M, Foulquie-Moreno MR, Thevelein JM. Looking beyond Saccharomyces: the potential of non-conventional yeast species for desirable traits in bioethanol fermentation. *FEMS Yeast Res* 2015;**15**.

Rasmusson AG, Geisler DA, Moller IM. The multiplicity of dehydrogenases in the electron transport chain of plant mitochondria. *Mitochondrion* 2008;**8**: 47-60.

Rasmusson AG, Soole KL, Elthon TE. Alternative NAD(P)H dehydrogenases of plant mitochondria. *Annu Rev Plant Biol* 2004;**55**: 23-39.

Ravin NV, Eldarov MA, Kadnikov VV, Beletsky AV, Schneider J, Mardanova ES, . . . Skryabin KG. Genome sequence and analysis of methylotrophic yeast *Hansenula polymorpha* DL1. *BMC Genomics* 2013;**14**: 837.

Rebnegger C, Vos T, Graf AB, Valli M, Pronk JT, Daran-Lapujade P, Mattanovich D. *Pichia pastoris* Exhibits High Viability and a Low Maintenance Energy Requirement at Near-Zero Specific Growth Rates. *Appl Environ Microbiol* 2016;**82**: 4570-83.

Riley R, Haridas S, Wolfe KH, Lopes MR, Hittinger CT, Goker M, . . . Jeffries TW. Comparative genomics of biotechnologically important yeasts. *Proc Natl Acad Sci U S A* 2016;**113**: 9882-7.

Robinson JT, Thorvaldsdottir H, Winckler W, Guttman M, Lander ES, Getz G, Mesirov JP. Integrative genomics viewer. *Nat Biotechnol* 2011;**29**: 24-6.

Robinson MD, McCarthy DJ, Smyth GK. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* 2010;**26**: 139-40.

Rocca-Serra P, Brandizi M, Maguire E, Sklyar N, Taylor C, Begley K, . . . Sansone SA. ISA software suite: supporting standards-compliant experimental annotation and enabling curation at the community level. *Bioinformatics* 2010;**26**: 2354-6.

Rogelj J, Shindell D, Jiang K, Fifita S, Forster P, Ginzburg V, . . . Zhou W. Chapter 2: Mitigation pathways compatible with 1.5°C in the context of sustainable development. In: Masson-Delmotte V, Zhai P, Pörtner H-O, Roberts D, Skea J, Shukla PR, Pirani A, Moufouma-Okia W, Péan C, Pidcock R, Connors S, Matthews JBR, Chen Y, Zhou X, Gomis MI, Lonnoy E, Maycock T, Tignor M, Waterfield T (eds.) *Global Warming of 1,5°C An IPCC special report on the impacts of global warming of 1,5°C above pre-industrial levels and related global greenhouse gas emission pathways, in the context of strengthening the global response to the threat of climate change, sustainable development, and efforts to eradicate poverty*: Intergovernmental Panel on Climate Change, 2018.

Roman H. A system selective for mutations affecting the synthesis of adenine in yeast. *Compt Rend Trav Lab Carlsberg, Ser physiol* 1956;**26**: 299-314.

Rothstein RJ. One-step gene disruption in yeast. *Methods Enzymol* 1983;**101**: 202-11.

Russell JB. The energy spilling reactions of bacteria and other organisms. *J Mol Microbiol Biotechnol* 2007;**13**: 1-11.

Ryabova OB, Chmil OM, Sibirny AA. Xylose and cellobiose fermentation to ethanol by the thermotolerant methylotrophic yeast *Hansenula polymorpha*. *FEMS Yeast Res* 2003;**4**: 157-64.

Ryan OW, Skerker JM, Maurer MJ, Li X, Tsai JC, Poddar S, . . . Cate JH. Selection of chromosomal DNA libraries using a multiplex CRISPR system. *eLife* 2014;**3**.

Salazar AN, Gorter de Vries AR, van den Broek M, Wijsman M, de la Torre Cortes P, Brickwedde A, . . . Abeel T. Nanopore sequencing enables near-complete de novo assembly of *Saccharomyces cerevisiae* reference strain CEN.PK113-7D. *FEMS Yeast Res* 2017;**17**: fox074.

Sapcariu SC, Kanashova T, Weindl D, Ghelfi J, Dittmar G, Hiller K. Simultaneous extraction of proteins and metabolites from cells in culture. *MethodsX* 2014;**1**: 74-80.

Saraya R, Gidijala L, Veenhuis M, van der Klei IJ. Tools for genetic engineering of the yeast *Hansenula polymorpha*. *Methods Mol Biol* 2014;**1152**: 43-62.

Saraya R, Krikken AM, Kiel JA, Baerends RJ, Veenhuis M, van der Klei IJ. Novel genetic tools for *Hansenula polymorpha*. *FEMS Yeast Res* 2012;**12**: 271-8.

Sauer B. Functional expression of the cre-lox site-specific recombination system in the yeast *Saccharomyces cerevisiae*. *Mol Cell Biol* 1987;**7**: 2087-96.

Sauer M, Porro D, Mattanovich D, Branduardi P. 16 years research on lactic acid production with yeast - ready for the market? *Biotechnol Genet Eng Rev* 2010;**27**: 229-56.

Saxena RK, Saran S, Isar J, Kaushik R. Production and Applications of Succinic Acid. In: Pandey A, Negi S, Soccol CR (eds.) *Current Developments in Biotechnology and Bioengineering*: Elsevier, 2017, 601-30.

Schägger H, Pfeiffer K. Supercomplexes in the respiratory chains of yeast and mammalian mitochondria. *EMBO J* 2000;**19**: 1777-83.

Schwann T. Vorläufige Mitteilung betreffend Versuche über die Weingärung und Fäulnis. *Ann Phys* 1837;**11**: 184.

Schwartz CM, Hussain MS, Blenner M, Wheeldon I. Synthetic RNA Polymerase III Promoters Facilitate High-Efficiency CRISPR-Cas9-Mediated Genome Editing in *Yarrowia lipolytica*. *ACS Synth Biol* 2016;**5**: 356-9.

Schwitzguebel JP, Palmer JM. Properties of mitochondria as a function of the growth stages of *Neurospora crassa*. *J Bacteriol* 1982;**149**: 612-9.

Shao S, Ren C, Liu Z, Bai Y, Chen Z, Wei Z, . . . Xu K. Enhancing CRISPR/Cas9-mediated homology-directed repair in mammalian cells by expressing *Saccharomyces cerevisiae* Rad52. *Int J Biochem Cell Biol* 2017;**92**: 43-52.

Shrivastav M, De Haro LP, Nickoloff JA. Regulation of DNA double-strand break repair pathway choice. *Cell Res* 2008;**18**: 134-47.

Sigrist CJ, Cerutti L, de Castro E, Langendijk-Genevaux PS, Bulliard V, Bairoch A, Hulo N. PROSITE, a protein domain database for functional characterization and annotation. *Nucleic Acids Res* 2010;**38**: D161-6.

Silverstein T. The mitochondrial phosphate-to-oxygen ratio is not an integer. *Biochem Mol Biol Educ* 2005;**33**: 416-7.

Singh MV, Weil PA. A method for plasmid purification directly from yeast. *Anal Biochem* 2002;**307**: 13-7.

Siso MIG. The biotechnological utilization of cheese whey: A review. *Bioresour Technol* 1996;**57**: 1-11.

Small WC, McAlister-Henn L. Identification of a cytosolically directed NADH dehydrogenase in mitochondria of *Saccharomyces cerevisiae*. *J Bacteriol* 1998;**180**: 4051-5.

Smith PK, Krohn RI, Hermanson GT, Mallia AK, Gartner FH, Provenzano MD, . . . Klenk DC. Measurement of protein using bicinchoninic acid. *Anal Biochem* 1985;**150**: 76-85.

Snoek IS, van der Krogt ZA, Touw H, Kerkman R, Pronk JT, Bovenberg RA, . . . Daran JM. Construction of an hdfA *Penicillium chrysogenum* strain impaired in non-homologous end-joining and analysis of its potential for functional analysis studies. *Fungal Genet Biol* 2009;**46**: 418-26.

Sonnleitner B, Chmiel H. Wachstum: Kinetik und Prozessführung. In: Chmiel H (ed.) *Bioprozesstechnik* volume 3. Heidelberg, 2011.

Souciet J, Aigle M, Artiguenave F, Blandin G, Bolotin-Fukuhara M, Bon E, . . . Weissenbach J. Genomic exploration of the hemiascomycetous yeasts: 1. A set of yeast species for molecular evolution studies. *FEBS Lett* 2000;**487**: 3-12.

Spohner SC, Schaum V, Quitmann H, Czermak P. *Kluyveromyces lactis*: An emerging tool in biotechnology. *J Biotechnol* 2016;**222**: 104-16.

Steinborn G, Boer E, Scholz A, Tag K, Kunze G, Gellissen G. Application of a wide-range yeast vector (CoMed) system to recombinant protein production in dimorphic *Arxula adeninivorans*, methylotrophic *Hansenula polymorpha* and other yeasts. *Microb Cell Fact* 2006;**5**: 33.

Stock D, Leslie AG, Walker JE. Molecular architecture of the rotary motor in ATP synthase. *Science* 1999;**286**: 1700-5.

Storici F, Durham CL, Gordenin DA, Resnick MA. Chromosomal site-specific double-strand breaks are efficiently targeted for repair by oligonucleotides in yeast. *Proc Natl Acad Sci U S A* 2003;**100**: 14994-9.

Storici F, Lewis LK, Resnick MA. In vivo site-directed mutagenesis using oligonucleotides. *Nat Biotechnol* 2001;**19**: 773-6.

Stovicek V, Holkenbrink C, Borodina I. CRISPR/Cas system for yeast genome engineering: advances and applications. *FEMS Yeast Res* 2017;**17**.

Stroud DA, Surgenor EE, Formosa LE, Reljic B, Frazier AE, Dibley MG, . . . Ryan MT. Accessory subunits are integral for assembly and function of human mitochondrial complex I. *Nature* 2016;**538**: 123-6.

Suh SO, Zhou JJ. Methylotrophic yeasts near *Ogataea* (*Hansenula*) *polymorpha*: a proposal of *Ogataea angusta* comb. nov. and *Candida parapolymorpha* sp. nov. *FEMS Yeast Res* 2010;**10**: 631-8.

Supek F, Bosnjak M, Skunca N, Smuc T. REVIGO summarizes and visualizes long lists of gene ontology terms. *PloS one* 2011;**6**: e21800.

Suwannarangsee S, Kim S, Kim OC, Oh DB, Seo JW, Kim CH, . . . Kwon O. Characterization of alcohol dehydrogenase 3 of the thermotolerant methylotrophic yeast *Hansenula polymorpha*. *Appl Microbiol Biotechnol* 2012;**96**: 697-709.

Suwannarangsee S, Oh DB, Seo JW, Kim CH, Rhee SK, Kang HA, . . . Kwon O. Characterization of alcohol dehydrogenase 1 of the thermotolerant methylotrophic yeast *Hansenula polymorpha*. *Appl Microbiol Biotechnol* 2010;**88**: 497-507.

Swiat MA, Dashko S, den Ridder M, Wijsman M, van der Oost J, Daran JM, Daran-Lapujade P. FnCpf1: a novel and efficient genome editing tool for *Saccharomyces cerevisiae*. *Nucleic Acids Res* 2017;**45**: 12585-98.

Symington LS. Role of *RAD52* epistasis group genes in homologous recombination and double-strand break repair. *Microbiol Mol Biol Rev* 2002;**66**: 630-70, table of contents.

Szostak JW, Orr-Weaver TL, Rothstein RJ, Stahl FW. The double-strand-break repair model for recombination. *Cell* 1983;**33**: 25-35.

Takahashi T, Masuda T, Koyama Y. Enhanced gene targeting frequency in *ku70* and *ku80* disruption mutants of *Aspergillus sojae* and *Aspergillus oryzae*. *Mol Genet Genomics* 2006;**275**: 460-70.

Tani Y, Yamada K. Diversity in Glycerol Metabolism of Methylotrophic Yeasts. *FEMS Microbiol Lett* 1987;**40**: 151-3.

Tappe W, Laverman A, Bohland M, Braster M, Rittershaus S, Groeneweg J, van Verseveld HW. Maintenance energy demand and starvation recovery dynamics of *Nitrosomonas europaea* and *Nitrobacter winogradskyi* cultivated in a retentostat with complete biomass retention. *Appl Environ Microbiol* 1999;**65**: 2471-7.

Tarrio N, Cerdan ME, Gonzalez Siso MI. Characterization of the second external alternative dehydrogenase from mitochondria of the respiratory yeast *Kluyveromyces lactis*. *Biochim Biophys Acta* 2006;**1757**: 1476-84.

Tarrio N, Diaz Prado S, Cerdan ME, Gonzalez Siso MI. The nuclear genes encoding the internal (*KlNDI1*) and external (*KlNDE1*) alternative NAD(P)H:ubiquinone oxidoreductases of mitochondria from *Kluyveromyces lactis*. *Biochim Biophys Acta* 2005;**1707**: 199-210.

Terentiev Y, Pico AH, Boer E, Wartmann T, Klabunde J, Breuer U, . . . Kunze G. A wide-range integrative yeast expression vector system based on *Arxula adeninivorans*-derived elements. *J Ind Microbiol Biotechnol* 2004;**31**: 223-8.

Toivola A, Yarrow D, van den Bosch E, van Dijken JP, Scheffers WA. Alcoholic Fermentation of d-Xylose by Yeasts. *Appl Environ Microbiol* 1984;**47**: 1221-3.

Tsakraklides V, Brevnova E, Stephanopoulos G, Shaw AJ. Improved Gene Targeting through Cell Cycle Synchronization. *PloS one* 2015;**10**: e0133434.

Ullmann F. *Bier. Enzyklopädie der technischen Chemie* volume 2. Berlin: Urban & Schwarzenberg, 1915.

van Bodegom P. Microbial maintenance: a critical review on its quantification. *Microb Ecol* 2007;**53**: 513-23.

van der Walt JP. New Combinations in the Genera *Brettanomyces*, *Kluyveromyces*, *Lodderomyces* and *Wingea*. *Bothalia* 1971;**10**: 417-8.

van Dijk R, Faber KN, Kiel JA, Veenhuis M, van der Klei I. The methylotrophic yeast *Hansenula polymorpha*: a versatile cell factory. *Enzyme Microb Technol* 2000;**26**: 793-800.

van Dijken JP, Bauer J, Brambilla L, Duboc P, Francois JM, Gancedo C, . . . Pronk JT. An interlaboratory comparison of physiological and genetic properties of four *Saccharomyces cerevisiae* strains. *Enzyme Microb Technol* 2000;**26**: 706-14.

van Dijken JP, Otto R, Harder W. Growth of *Hansenula polymorpha* in a methanol-limited chemostat. Physiological responses due to the involvement of methanol oxidase as a key enzyme in methanol metabolism. *Arch Microbiol* 1976;**111**: 137-44.

van Gulik WM, Antoniewicz MR, deLaat WT, Vinke JL, Heijnen JJ. Energetics of growth and penicillin production in a high-producing strain of *Penicillium chrysogenum*. *Biotechnol Bioeng* 2001;**72**: 185-93.

van Gulik WM, Heijnen JJ. A metabolic network stoichiometry analysis of microbial growth and product formation. *Biotechnol Bioeng* 1995;**48**: 681-98.

Van Urk H, Bruinenberg PM, Veenhuis M, Scheffers WA, Van Dijken JP. Respiratory capacities of mitochondria of *Saccharomyces cerevisiae* CBS 8066 and *Candida utilis* CBS 621 grown under glucose limitation. *Antonie van Leeuwenhoek* 1989;**56**: 211-20.

van Zutphen T, Baerends RJ, Susanna KA, de Jong A, Kuipers OP, Veenhuis M, van der Klei IJ. Adaptation of *Hansenula polymorpha* to methanol: a transcriptome analysis. *BMC Genomics* 2010;**11**: 1.

Vanrolleghem PA, de Jong-Gubbels P, van Gulik WM, Pronk JT, van Dijken JP, Heijnen S. Validation of a metabolic network for Saccharomyces cerevisiae using mixed substrate studies. *Biotechnol Prog* 1996;**12**: 434-48.

VDI. Biotechnologie - Vernetzung von Naturwissenschaft mit Ingenieur-Know-how. *VDI-Fachbereich Biotechnologie* 2015.

Veiga A, Arrabaca JD, Loureiro-Dias MC. Cyanide-resistant respiration, a very frequent metabolic pathway in yeasts. *FEMS Yeast Res* 2003;**3**: 239-45.

Verduyn C. Physiology of yeasts in relation to biomass yields. *Antonie van Leeuwenhoek* 1991;**60**: 325-53.

Verduyn C, Postma E, Scheffers WA, van Dijken JP. Physiology of *Saccharomyces cerevisiae* in anaerobic glucose-limited chemostat cultures. *J Gen Microbiol* 1990;**136**: 395-403.

Verduyn C, Postma E, Scheffers WA, Van Dijken JP. Effect of benzoic acid on metabolic fluxes in yeasts: a continuous-culture study on the regulation of respiration and alcoholic fermentation. *Yeast* 1992;**8**: 501-17.

Verduyn C, Stouthamer AH, Scheffers WA, van Dijken JP. A theoretical evaluation of growth yields of yeasts. *Antonie van Leeuwenhoek* 1991;**59**: 49-63.

Vieira Gomes AM, Souza Carmo T, Silva Carvalho L, Mendonca Bahia F, Parachin NS. Comparison of Yeasts as Hosts for Recombinant Protein Production. *Microorganisms* 2018;**6**: 38.

Villadsen J, Nielsen JH, Lidén G, Nielsen JH. *Bioreaction engineering principles*: 3rd Edition. New York: Springer, 2011.

Vinothkumar KR, Montgomery MG, Liu S, Walker JE. Structure of the mitochondrial ATP synthase from *Pichia angusta* determined by electron cryo-microscopy. *Proc Natl Acad Sci U S A* 2016;**113**: 12709-14.

von Jagow G, Klingenberg M. Pathways of hydrogen in mitochondria of *Saccharomyces carlsbergensis*. *Eur J Biochem* 1970;**12**: 583-92.

Vos T, Hakkaart XD, de Hulster EA, van Maris AJ, Pronk JT, Daran-Lapujade P. Maintenance-energy requirements and robustness of *Saccharomyces cerevisiae* at aerobic near-zero specific growth rates. *Microb Cell Fact* 2016;**15**: 111.

Voulgaris I, O'Donnell A, Harvey LM, McNeil B. Inactivating alternative NADH dehydrogenases: enhancing fungal bioprocesses by improving growth and biomass yield? *Sci Rep* 2012;**2**: 322.

Wach A, Brachat A, Pohlmann R, Philippsen P. New heterologous modules for classical or PCR-based gene disruptions in *Saccharomyces cerevisiae*. *Yeast* 1994;**10**: 1793-808.

Wagner JM, Alper HS. Synthetic biology and molecular genetics in non-conventional yeasts: Current tools and future advances. *Fungal Genet Biol* 2016;**89**: 126-36.

Wallrath J, Schmidt M, Weiss H. Concomitant loss of respiratory chain NADH:ubiquinone reductase (complex I) and citric acid accumulation in *Aspergillus niger*. *Appl Microbiol Biotechnol* 1991;**36**: 76-81.

Wang W-H, Himeda Y, Muckerman JT, Manbeck GF, Fujita E. CO2 Hydrogenation to Formate and Methanol as an Alternative to Photo- and Electrochemical CO2 Reduction. *Chemical Reviews* 2015;**115**: 12936-73.

Waterhouse AM, Procter JB, Martin DM, Clamp M, Barton GJ. Jalview Version 2--a multiple sequence alignment editor and analysis workbench. *Bioinformatics* 2009;**25**: 1189-91.

Watt IN, Montgomery MG, Runswick MJ, Leslie AG, Walker JE. Bioenergetic cost of making an adenosine triphosphate molecule in animal mitochondria. *Proc Natl Acad Sci U S A* 2010;**107**: 16823-7.

Weiss B, Richardson CC. Enzymatic breakage and joining of deoxyribonucleic acid, I. Repair of single-strand breaks in DNA by an enzyme system from *Escherichia coli* infected with T4 bacteriophage. *Proc Natl Acad Sci U S A* 1967;**57**: 1021-8.

Weninger A, Hatzl AM, Schmid C, Vogl T, Glieder A. Combinatorial optimization of CRISPR/Cas9 expression enables precision genome engineering in the methylotrophic yeast *Pichia pastoris*. *J Biotechnol* 2016;**235**: 139-49.

Weusthuis RA, Visser W, Pronk JT, Scheffers WA, van Dijken JP. Effects of oxygen limitation on sugar metabolism in yeasts: a continuous-culture study of the Kluyver effect. *Microbiology* 1994;**140 (Pt 4)**: 703-15.

Wilkinson MD, Dumontier M, Aalbersberg IJ, Appleton G, Axton M, Baak A, . . . Mons B. The FAIR guiding principles for scientific data management and stewardship. *Sci Data* 2016;**3**: 160018.

Xiao W, Wang RS, Handy DE, Loscalzo J. NAD(H) and NADP(H) Redox Couples and Cellular Energy Metabolism. *Antioxid Redox Signal* 2018;**28**: 251-72.

Yamada Y, Maeda K, Mikata K. The phylogenetic relationships of the hat-shaped ascospore-forming, nitrate-assimilating *Pichia* species, formerly classified in the genus *Hansenula* Sydow et Sydow, based on the partial sequences of 18S and 26S ribosomal RNAs (Saccharomycetaceae): the proposals of three new genera, *Ogataea*, *Kuraishia*, and *Nakazawaea*. *Biosci Biotechnol Biochem* 1994;**58**: 1245-57.

Yang Z, Edwards H, Xu P. CRISPR-Cas12a/Cpf1-assisted precise, efficient and multiplexed genome-editing in Yarrowia lipolytica. *Metab Eng Commun* 2020;**10**: e00112.

Yurimoto H, Oku M, Sakai Y. Yeast methylotrophy: metabolism, gene regulation and peroxisome homeostasis. *Int J Microbiol* 2011;**2011**: 101298.

Zhu Z, Hu Y, Teixeira PG, Pereira R, Chen Y, Siewers V, Nielsen J. Multidimensional engineering of *Saccharomyces cerevisiae* for efficient synthesis of medium-chain fatty acids. *Nature Catalysis* 2020;**3**: 64-74.

Zimmerman SB, Little JW, Oshinsky CK, Gellert M. Enzymatic joining of DNA strands: a novel reaction of diphosphopyridine nucleotide. *Proc Natl Acad Sci U S A* 1967;**57**: 1841-8.

Acknowledgements

This thesis is the result of an incredibly fun, fulfilling, and sometimes stressful journey filled with many lessons on setting priories, making decisions and growing as a person and as a scientist. This journey was not taken alone, and many people contributed to its success. Only now that I look back at this time with quite some distance, during July 2021, I realize how shaping it really was. Here, I want to express my gratitude to you.

First, my three daily supervisors, in chronological order: **Ton**, **Jack** and **Robert**. Thank you all for the support, guidance, enthusiasm, trust and freedom. **Ton**, thank you for setting up this project and for taking me on board at IMB. I very much enjoyed your supervision at the beginning of the project, always being extra sharp and critical in a constructive way. No matter how well I tried to design an experiment or write a report, you always found something to improve and discussing these things with you was a great experience. I still start to sweat when I think back to my interview with you. Partly because it was one of the hottest days of 2015, I came dressed up in a suit, and we talked for hours outside in the botanical garden (you in shorts and a t-shirt). And partly because of your interrogation-like interview conduction skills that foreshadowed how critical and detail-oriented my supervisor-to-be would be like. I loved it. I also had the privilege to teach together with you, and witnessing your passion for trying to get every last student understand the concepts was great. I'm happy that you found your place in Sweden. **Jack**, you were the one constant in my PhD project. Despite an increasing number of responsibilities, you always found the time to give advice and never tuned down your curiosity for the project (mine, or anyone else's). Your seemingly unlimited enthusiasm for any aspect of science was a great source of new ideas and motivation (if they would let you, you would probably add five additional research lines to every PhD project). Also, your leadership style of integrity and humility as section leader of IMB during my time there continues to be an inspiration to this day. I admire the kind of environment and atmosphere you have built together with the other staff members that I got to enjoy during my project. So, Jack, how *did* you get here? **Robert**, you took over the project as you transitioned from PhD to fresh PI and from day one, I knew there were not going to be any problems. Thank you for giving me so much freedom in pursuing the PhD, yet still providing guidance whenever needed. You always kept things fun and enjoyable, were very understanding and approachable, and never hesitated to jump in and help when required. Whenever we discussed any aspect of the project you helped me focus and reflect, came up with great solutions, and just talking to you usually made the majority of problems disappear. I am very curious where your academic (or non-academic) career will lead you to, and I will make sure to follow your achievements.

Last time I checked Google Scholar, you were well ahead of Robert Mans #2 in number of citations - first milestone achieved.

A special thank you also to you, **Jean-Marc**. You were my joker supervisor. When I found you in your office to get insights and suggestions on molecular biology or anything else, I usually came out with a good idea or new view on what to do next. You were essential for the genetic engineering work early in my PhD project ("WE GOT RED COLONIES!!!"), and many things wouldn't have been possible without you.

Collaboration is where the magic happens, and besides my supervisors, I got the opportunity to work with great people on every research line that ended up in this thesis. Thanks go to **Arthur** and **Thomas**, as well as **Javier Varela** and **John Morrissey** for the great "yeasty boys" collaboration and teamwork on Chapter 2. Thank you **Matthijs** for spending a lot of time on transferring IMB fermentation processes to Riffyn for Chapter 3, as well as **Timothy Gardner**, **Laura Jennings-Antipov** and **Jack Morel** for the collaboration on a somewhat atypical project for IMB. This project was tough to push forward and the manuscript was completely rewritten several times. Without me realizing it at the time, this work was likely the most important one for me for the following years to come. Only now outside of the academic environment, where demonstrating something once and getting it published does not equal success anymore, and the number of people that need to access the same datasets is much bigger, I begin to truly understand the need for such rigor in carefully collecting, structuring, analyzing and sharing data. Chapter 4 was a great collaboration and team effort, **Xavier**. I very much enjoyed us being *Team Daddy*, and I am glad you finally got to work with someone to make all the jokes you never could before. Thanks also go to **Pascale** for the input and discussions on the omics part of this work, as well as to **André Vente** for processing the proteomics samples. Chapter 5 was a great mix of expertise from two Biotechnology sections. **Albert** and **Duncan**, it was always energizing to talk to you about respiration and the amazing molecular protein machines that make it happen, and to see how passionate you are about this topic. It was great to get an "outside IMB" perspective from time to time that focused much more on the biochemistry itself. A big thank you also to **Jolanda**, your super-organized help with strain construction gave this project a big boost. Finally, I want to thank **Kirsten Benjamin** and **Liang Wu**, who, from an industry point of view gave valuable input throughout the entire PhD project.

From the perspective of a PhD student, IMB is a well-oiled science machine and high standards are kept up by the staff and project technicians. *Meten heet weten*, but good methods are everything. Your importance for IMB can barely be overstated. Thank you all for being so devoted, yet so approachable. Most of the time, help was available almost immediately. **Erik**, thank you for staying on top of the fermentation lab and creating such a well-organized working experience for all of us. IMB is ready to go next

level with the all-connected and -controllable setup. **Matthijs**, (word I cannot say here), thank you for teaching me "proper fermentation, Delft style" in my first year. I still have the German discounter whiskey you made me buy once, and I will keep it until our paths cross again. **Marijke**, thank you for being so compassionate, and for all the help in Mickel's lab even late at night. You showed me that, *yes we can* isolate mitochondria from *Ogataea*. It was great working with you and to see your creativity unfold in the lab, for example when you built contraptions to mix samples *juuust right* for good resuspension. You have a way of approaching problems that makes hard things seem easy. **Pilar**, more than anything I miss our good talks inside and outside the labs, between samples or over cake. Rocking Wesley's defense as paranymph buddies was a lot of fun. Big hug to Spain. **Marcel**, thank you for all the help with sequencing, genome assembly, transcriptomics, … the list goes on. It was always very interesting to peek over your shoulder and listening to your explanation when you were working on the command line. I'm sure, one day, your dream will come true and you will get to meet Kevin Verstrepen.

I could not imagine a better environment than the IMB section to pursue a PhD. It is a constantly changing, but always nice and welcoming environment. Ultimately, a section is just the people that it consists of, and I experienced IMB as filled with unique people dedicated to science, pushing each other to do better with minimal drama or negativity. Quick passes by the coffee corner to discuss an idea regularly turned into long conversations or "science dinners" later that day. Many victories were shared, and countless problems were troubleshot together. We had what today feels like an endless number of distractions in the form of parties, conference visits (=parties with a day programme), dinners, trips, group events, movie & gaming nights, … thank you, **all of you**, for contributing to this great environment. I am missing many of you. While to make a complete list here would be too much, I would like to mention some groups and people in particular. **The Energetics Group.** A great forum for anyone to discuss results and (crazy) ideas and get instant feedback, which I was always looking forward to. From early editions together with Ton, Robert, Wesley and Nicolò (usually 5 guys in uniform look: jeans and a black t-shirt), till more recent editions with intermittent visits from Jack, and also Sophie and Aafke joining – Energetics was always a great place to laugh, especially about Nicolò's enzymes of certain origin. I remember several occasions of Jack joining the meeting slightly late only to find people teared up from laughing, holding their sides and having to postpone the meeting for some time (Jack, your jokes didn't help either). **The Jam Sessions Group.** Mario, Raúl, Chema, Sanne and the various "guest stars" along the way … we were proper *Kleinkünstler*. I hugely enjoyed our sometimes weekly, sometimes less frequent jams at the sports center. It was a great creative outlet, and having just listened to some of the recordings again (I think we called it *Space Vikings*) I have to say it wasn't even half bad!!! **The International's**

Dinner Club. Not an exclusive club and often with guests, but definitely with a bunch of regulars: Francine, Michal, Nicolò, Pilar, Sofia, Wesley. Our regular dinners were always a joy and a welcome opportunity to shut down from work (and also a great way of learning the best restaurants in town). **Ángel**, thank you for being such a welcoming office mate in my first months (still back in the Julianalaan dungeons), and for explaining a lot about the group to me. **Arthur**, thank you for our pizza nights that never went as planned, for teaching me how to vortex like a pro (I am passing on the knowledge and people are amazed every time), and especially for keeping an eye out for the well-being of my students during my first year when I wasn't very experienced yet at supervising. Teaming up as paranymphs for Ioannis defense with you was also great fun. **Aurin**, I very much enjoyed our discussions, be it any flavor of science, statistics or some absurd fictional thing you just came up with (which happened about daily). **Nicolò**, when you think about something you don't remove the box, there is just no box to begin with. Thank you for the countless fresh and valuable ideas you provided over the years, which this PhD project has benefitted from more than once. I consider anyone lucky that gets to work with you. Thank you also for being a regular partner in crime for working late, be it to still run those enzyme assays or because of intended or unintended 0.0****-you growth rates. I also enjoyed our collaboration on *Team Baldi and Baldy*, and I still have the mug. Being around you was fun, there was never a lack of memes, and over the years I even grew accustomed to your sense of humor (you still have to work on decreasing the eye-rolling factor, though). More *Strange Planet* comics will come your way, I promise. **Jasmijn**, having you as an office mate and generally around the group was amazing because you radiate energy like few other people. I hope you forgive me all the pranks that (mostly Robert and) I played on you, and I think you suffered the most from me blasting the 10-h version of *They're taking the Hobbits to Isengard* during every week's lab cleaning, but I regret nothing and would do it the same way again. **Sanne**, I miss having such a fun office mate as you, especially our *Friday-afternoon-share-the-stupidest/funniest-video-you-could-find* tradition that often drew a small crowd. Also, thank you for promoting me from train-buddy to car-buddy during our time commuting from Leiden. **Xavier,** just like your typical student victim, I equally benefitted from your "instant explanation mode". This is what many years of LS&T/IMB experience coupled with a love of teaching looks like, and was extra helpful as you introduced me into the world of near-zero growth. Thank you for all the great explanations and project ideas over the years, and general good talks and advice also after we both already left the group. *Dein Deutsch ist toll und wir können gerne öfters plaudern*. Thank you, **Ioannis** and **Wesley**, for entrusting me with being your paranymph, and thank you **Jasmijn** and **Mario**, for organizing my goodbye dinner and such a nice present when I left IMB. Finally, **Michal & Wesley**, not many words are

needed here. Thank you for all the silly fun we had and have, for keeping the bromance alive, and for being my paranymphs. The legend of the t-shirt will be continued.

I was fortunate to have a group of wonderful students joining me on this journey. **Veronica**, **Nikola**, **Siem**, **Lisan**, **Jildau**, **Effie**, **Thijs**, **Álvaro**, **Janine**. You did a huge bulk of the experimental work that is in this thesis, and many of you have made it onto the publications that form the chapters of this thesis – thank you!! I am very proud of you because you all showed a clear development from beginning to end of your projects, and it is very exciting to see that some of you started your own PhDs and publish your own research now.

Apilena, **Astrid** and **Jannie**, if it wasn't for your work, the research output of the Biotechnology department would be a lot lower. Having access to consistently prepared media, bioreactors and well-maintained sensors over the entire duration of the PhD project is almost invaluable. **Marcel**, thank you for checking and maintaining our fermentation equipment so well, and on short notice, especially just before those long, delicate retentostat experiments.

A special thank you also goes to my Dutch roommates in Delfgauw at DSW 228 and 230, **Eric**, **Floris**, **Kristan**, **Niels**, **Roos**, **Ruben**, **Sem**, **Stephan** and **Willem**. You welcomed me with open arms when I just arrived in a new country, and fully included me in what I would describe as the "Dutch student" way of living: beer tastings, theme parties, not following the cleaning plan and a general carefree attitude. Great times.

Thank you, **Mama und Papa**. I cannot pinpoint when this happened, but you have instilled in me the desire to learn and understand the world around me. This is for me the greatest gift, as I will never get bored and enjoy learning for the rest of my life. Thank you for always supporting me and giving me the freedom to pursue what I wanted.

Finally, thank you **Laura**. You accompanied me for the entire duration of this journey. I distinctly remember how you encouraged me before the interview on that hot summer day in 2015, and right now you are cracking the metaphorical whip (and threatening to buy a real one) so I finally finish this book even though my head is revolving around completely different projects. We celebrated the successes together, and you were always there and understanding during the more stressful times. I do not know a person that is more compassionate and empathic than you, and I hope that you will be able to always retain this quality of yours. I am looking forward to our future adventures.

Curriculum vitae

Hannes Jürgens was born on May 26th, 1988 in Achim, Germany. After finishing pre-university education in 2008, he enrolled in a Biotechnology and Process Engineering bachelor's program at Flensburg University of Applied Sciences, Germany. Hannes performed his bachelor's research project at the company e-nema in Kiel under the supervision of Dr. Arne Peters. In this project he worked on development of a lab-scale fermentation process that utilized *Pseudomonas putida* as catalyst for the biotransformation of the organic waste compound limonene into higher value compounds.

After finishing his bachelor's studies, Hannes continued his stay at e-nema for 1.5 years to further develop and scale the biotransformation process as a Research Assistant. Afterwards, he enrolled in the Biotechnology master's program at Aachen University of Applied Sciences in Jülich, Germany, where he followed a wide selection of courses covering topics from industrial, medical and agricultural biotechnology. For his master's research project, Hannes joined the genetics department of DSM in Delft, the Netherlands, for 6 months, where he worked on screening and engineering of non-ribosomal peptide synthetases under the supervision of Rémon Boer and Dr. Ulrike Müller.

After graduating in 2015, Hannes joined the Industrial Microbiology section (IMB) at Delft University of Technology to start his PhD project under the supervision of (in various combinations) Prof.dr.ir. Ton van Maris, Prof.dr. Jack Pronk and Dr.ir. Robert Mans. This project, the results of which are described in this thesis, focused on investigation and engineering of respiratory energy-coupling in yeasts, with special focus on the thermotolerant *Ogataea* genus. In April 2020, Hannes joined the startup company LenioBio in Germany, where he works on scaling a versatile eukaryotic cell-free protein expression system.

List of publications

Dekker WJC, **Juergens H,** Ortiz-Merino RA, Mooiman C, van den Berg R, Kaljouw A, Mans R and Pronk JT. Respiratory reoxidation of NADH is a key contributor to high oxygen requirements of oxygen-limited cultures of *Ogataea parapolymorpha*. FEMS Yeast Research 22(1), 2022.

Juergens H, Mielgo-Gómez Á, Godoy-Hernández A, ter Horst J, Nijenhuis JM, McMillan DGG and Mans R. Physiological relevance, localization and substrate specificity of the alternative (type II) mitochondrial NADH dehydrogenases of *Ogataea parapolymorpha*. Biorxiv, 2021.

Juergens H*, Hakkaart XDV*, Bras JE, Vente A, Wu L, Benjamin KR, Pronk JT, 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 86(15), 2020; **joint first authorship*.

Rajkumar AS, Varela JA, **Juergens H,** Daran J-MG and Morrissey JP. Biological parts for *Kluyveromyces marxianus* synthetic biology. Frontiers in Bioengineering and Biotechnology 7, 2019.

Juergens H*, Niemeijer M*, Jennings-Antipov LD*, Mans R, Morel J, van Maris AJA, Pronk JT and Gardner TS. Evaluation of a novel cloud-based software platform for structured experiment design and linked data analytics. Scientific Data 5, 2018; **joint first authorship*.

Juergens H, Varela JA, Gorter de Vries AR, Perli T, Gast VJM, Gyurchev NY, Rajkumar AS, Mans R, Pronk JT, Morrissey JP and Daran J-MG. Genome editing in *Kluyveromyces* and *Ogataea* yeasts using a broad-host-range Cas9/gRNA co-expression plasmid. FEMS Yeast Research 18(3), 2018.