Gene 492 (2012) 177–185

Contents lists available at SciVerse ScienceDirect

Gene

journal homepage: www.elsevier.com/locate/gene

Characterisation of the gene cluster for L-rhamnose catabolism in the yeast *Scheffersomyces* (*Pichia*) *stipitis*

Outi M. Koivistoinen [a], Mikko Arvas [a], Jennifer R. Headman [b], Martina Andberg [a], Merja Penttilä [a], Thomas W. Jeffries [b,c], Peter Richard [a,*]

[a] VTT Technical Research Centre of Finland, Espoo, Finland
[b] University of Wisconsin, Department of Bacteriology, WI 53706, USA
[c] USDA, Forest Products Laboratory, Madison 53726, USA

ARTICLE INFO

Article history:
Accepted 11 October 2011
Available online 21 October 2011

Received by Meghan Jendrysik

Keywords:
L-rhamnose catabolism
Scheffersomyces stipitis
L-rhamnonate dehydratase

ABSTRACT

In *Scheffersomyces* (*Pichia*) *stipitis* and related fungal species the genes for L-rhamnose catabolism *RHA1*, *LRA2*, *LRA3* and *LRA4* but not *LADH* are clustered. We find that located next to the cluster is a transcription factor, *TRC1*, which is conserved among related species. Our transcriptome analysis shows that all the catabolic genes and all genes of the cluster are up-regulated on L-rhamnose. Among genes that were also up-regulated on L-rhamnose were two transcription factors including the *TRC1*. In addition, in 16 out of the 32 analysed fungal species only *RHA1*, *LRA2* and *LRA3* are physically clustered. The clustering of *RHA1*, *LRA3* and *TRC1* is also conserved in species not closely related to *S. stipitis*. Since the *LRA4* is often not part of the cluster and it has several paralogues in L-rhamnose utilising yeasts we analysed the function of one of the paralogues, *LRA41* by heterologous expression and biochemical characterization. Lra41p has similar catalytic properties as the *Lra4p* but the transcript was not up-regulated on L-rhamnose. The *RHA1*, *LRA2*, *LRA4* and *LADH* genes were previously characterised in *S. stipitis*. We expressed the L-rhamnonate dehydratase, Lra3p, in *Saccharomyces cerevisiae*, estimated the kinetic constants of the protein and showed that it indeed has activity with L-rhamnonate.

© 2011 Elsevier B.V. All rights reserved.

1. Introduction

L-Rhamnose (L-6-deoxy-mannose) is a natural sugar found in hemicellulose and pectin. Numerous microorganisms use α-L-rhamnosidase to release L-rhamnose as a source for carbon and energy (Twerdochlib et al., 1994). Beyond hydrolysis, however, two different pathways for the catabolism of L-rhamnose are used: one that involves phosphorylated intermediates in which an isomerase is the first enzyme and one that uses non-phosphorylated intermediates with an oxidative pathway. The isomerase pathway has been described only in bacteria. It includes L-rhamnose isomerase (EC 5.3.1.14) (Wilson and Ajl, 1957a; Takagi and Sawada, 1964a), L-rhamnulokinase (EC 2.7.1.5) (Wilson and Ajl, 1957b; Takagi and Sawada, 1964b), and L-rhamnulose 1-phosphate aldolase (EC 4.1.2.19) (Sawada and Takagi, 1964). The metabolites are L-rhamnulose, L-rhamnulose-1-phosphate, dihydroxyacetone phosphate and L-lactaldehyde. L-lactaldehyde can then be reduced to 1,2-propendiol or oxidised to lactate by lactaldehyde reductase (EC 1.1.1.77) or lactaldehyde dehydrogenase (EC 1.2.1.22) respectively depending on the redox conditions (Baldoma and Aguilar, 1988). Gene sequences for all these enzymes have been described (Moralejo et al., 1993).

The oxidative pathway is present in bacteria and in fungi (Watanabe et al., 2008b) (Fig. 1). By this route L-rhamnose is first oxidised to L-rhamnono-γ-lactone by an NAD-utilising L-rhamnose-1-dehydrogenase. This inducible enzyme activity has been described in various yeast species including *Scheffersomyces stipitis*, *Aureobasidium* (*Pullularia*) *pullulans* and *Debaryomyces polymorphus* (Rigo et al., 1976; Vieira et al., 1979; Twerdochlib et al., 1994). The corresponding gene, *RHA1* (ABN68405), was cloned from *S. stipitis*; the active protein was expressed in a heterologous host, and the histidine-tagged protein was characterised. It was also shown that transcription is induced on L-rhamnose but not on other carbon sources (Koivistoinen et al., 2008).

The other enzymes of this pathway are L-rhamnono-γ-lactonase (EC 3.1.1.65), L-rhamnonate dehydratase (EC 4.2.1.90), 3,6-dideoxy-L-erythro-hexulosonic acid aldolase (L-2-keto-3-deoxy-rhamnonate aldolase) (EC 4.2.1.-) and L-lactaldehyde dehydrogenase (EC 1.2.1.22). The activities of these enzymes were described in different yeast species (Rigo et al., 1985; Twerdochlib et al., 1994), except for the lactonase which was suggested to be not necessary for a functional pathway since the lactone can hydrolyse spontaneously (Twerdochlib et al., 1994). The genes coding for the dehydrogenase, *RHA1*, lactonase, *LRA2* (ABN68602), a putative dehydratase, *LRA3* (ABN68404), and aldolase,

Abbreviations: TCA, Trichloroacetic acid; TBA, Thiobarbituric acid; ML, Maximum likelihood.

* Corresponding author. Tel.: +358 20 722 7190; fax: +358 20 722 7071.
E-mail address: Peter.Richard@vtt.fi (P. Richard).

0378-1119/$ – see front matter © 2011 Elsevier B.V. All rights reserved.
doi:10.1016/j.gene.2011.10.031

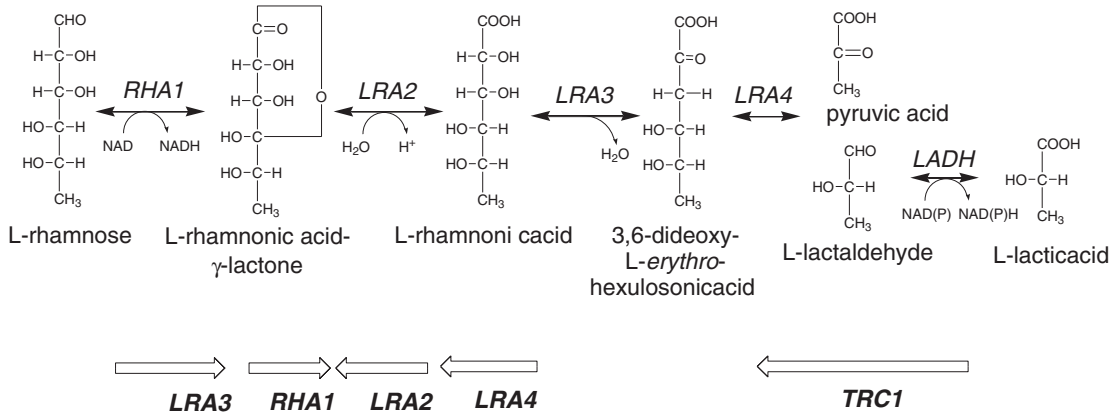

Fig. 1. Upper part: The eukaryotic L-rhamnose pathway. L-Rhamnose is converted to pyruvic acid and L-lactic acid using an oxidative pathway without phosphorylated intermediates similar to the Entner–Doudoroff pathway. The genes coding for the first four enzymes of this pathway are organised in a cluster whilst the LADH is not physically linked. Lower part: Orientation of the genes in the gene cluster.

LRA4 (ABN68603), were identified in *S. stipitis* and shown to be organised in a gene cluster (Watanabe et al., 2008b). Dehydrogenase (*RHA1*), lactonase (*LRA2*) and aldolase (*LRA4*) were heterologously expressed and the histidine-tagged proteins characterised, whereas the heterologous expression of the dehydratase did not result in an active enzyme (Watanabe et al., 2008b). In *S. stipitis* L-lactaldehyde is oxidised to L-lactate. The enzyme and the corresponding gene, LADH (ABN68636), were identified, the histidine-tagged protein characterised and the transcription was shown to be induced on L-rhamnose (Watanabe et al., 2008a).

In *S. stipitis* the four genes of the oxidative L-rhamnose pathway are situated side-by-side on chromosome 8, i.e. in a chromosomal gene cluster. However, for eukaryotes the definition of chromosomal gene clusters is loose (Lee and Sonnhammer, 2003) and does not require strict adjacency as demonstrated by *Trichoderma reesei* cellulase clusters of over 200 kb (Martinez et al., 2008) and the intervening genes of galactose utilisation clusters (Martchenko et al., 2007; Slot and Rokas, 2010). Based on the genome of *S. stipitis*, it has, in addition to L-rhamnose metabolism, numerous clusters for the metabolism of urea, cellobiose, maltose, iron metabolism, galactose metabolism, pyrimidine metabolism and dityrosine metabolism (Jeffries and Van Vleet, 2009). Functional gene clusters are found in fungal synthesis pathways of secondary metabolites such as antibiotics, toxins and pigments (Keller et al., 2005). In addition to L-rhamnose, fungal clusters of nutrient assimilation pathway genes have been identified at least for quinate (Giles et al., 1991), ethanol (Fillinger and Felenbok, 1996), proline (Hull et al., 1989), nitrate (Johnstone et al., 1990), inulin (Yuan et al., 2008) galactose (Hittinger et al., 2004; Slot and Rokas, 2010) allantoin (Cooper, 1996) and amylolytic pathway (Gomi et al., 2000). In addition, clusters of genes involved in pathogenesis and cellulose hydrolysis have been proposed (Covert et al., 1992; Martinez et al., 2008). Broad multigenome analyses of distribution and evolution of these clusters, however have been only carried out for nitrate (Slot and Hibbett, 2007) and allantoin (Cooper, 1996), and D-galactose clusters (Hittinger et al., 2004; Slot and Rokas, 2010).

Often published fungal metabolic gene clusters include an embedded pathway specific transcription factor with the Zn(II)2Cys6 zinc binuclear cluster-domain (InterPro: IPR001138) (Keller et al., 2005). For example, two pairs of maltose permeases and alpha-glucosidases in *S. stipitis* are flanked by Zn(II)2Cys6 genes, and a similar arrangement is found in *Saccharomyces cerevisiae*. Of the 9 clusters listed above, L-rhamnose, nitrate, D-galactose and allantoin have not been reported to contain a Zn(II)2Cys6 gene. The Zn(II)2Cys6 domain typically occurs together with the 'Fungal specific transcription factor' -domain (IPR007219) (Arvas et al., 2007). In genomes of Ascomycota subphyla Pezizomycotina the Zn(II)2Cys6 genes are about three times more abundant than in Saccharomycotina whilst the Pezizomycotina have only two times more genes than Saccharomycotina (significant difference between the subphyla with a p-value$<$0.05 in Wilcoxon rank sum test) (Arvas et al., 2007). The chromosomal co-localisation of Zn(II)2Cys6 with metabolic clusters and expansion in Pezizomycotina, known for their varied metabolism, could mean that Zn(II)2Cys6 are common regulators of metabolic clusters in Fungi.

In the present communication we analysed the distribution of the L-rhamnose cluster in 32 fungal genomes and found that the cluster in its entirety is conserved only in close relatives of *S. stipitis* i.e. *in Debaryomyces hansenii*, *Candida lusitaniae* and *Candida guilliermondii*. However, in these species an orthologous transcription factor with Zn(II)2Cys6 (IPR001138) and IPR007219 domains is always adjacent to the cluster and this transcription factor and *RHA1* and *LRA3* are found in close proximity i.e. in a chromosomal gene cluster in 16 of the 32 species including the basidiomycete *Cryptococcus neoformans*.

We also show that the transcription of all genes of the pathway is upregulated on L-rhamnose and that an active L-rhamnonate dehydratase can be expressed heterologously in *S. cerevisiae*. We further demonstrate that there is another aldolase in the *S. stipitis* genome that is located outside the cluster that has similar activities and substrate specificities but that is not upregulated on L-rhamnose.

2. Results

2.1. Presence of L-rhamnose catabolism genes in fungi

In order to study the distribution of L-rhamnose genes in Fungi we looked for them in 32 fungal genomes and compared how they were clustered.

Individual genes of the L-rhamnose cluster were found in Basidiomycota and Ascomycota (Table 1). Pezizomycotina, excluding *Coccidioides immitis* that has only LADH, have several (5 species, mainly Eurotiomycetes) or all genes (7 species, mainly Sordariomycetes) of the L-rhamnose catabolism cluster. All Eurotiomycetes lack the LRA4. However, many other undescribed members of the protein family dihydrodipicolinate synthetase (IPR002220), to which *LRA4* belongs are found in Eurotiomycetes and sequence analysis alone cannot assign an exact molecular function to them (data not shown). In contrast duplicates of *LRA4* are found in Saccharomycotina species.

A stronger division than in Pezizomycotina, is seen in Saccharomycotina, where 3 closely related species (*C. lusitaniae*, *D. hansenii* and *S. stipitis*) have all the genes of the cluster and 7 species none

Table 1
Counts of L-rhamnose catabolism genes in the species studied. Column "Cluster" specifies if a full (F) rhamnose cluster (*RHA1*, *LRA2*, *LRA3*, *LRA4* and *TRC1*) or just the core (C) cluster (*RHA1*, *LRA2* and *TRC1*) was found from a species (Fig. 2) and column 'Growth on L-rhamnose' whether, based on literature (Kurtzman and Fell, 1998) or CBS database (http://www.cbs.knaw.nl/databases/), the species is able to grow on rhamnose. Plus (+) sign indicates growth, minus (−) sign no growth and (w) as weak growth.

Phylum	Subphylum	Class	Name	Growth on L-rhamnose	Cluster	RHA1	LRA2	LRA3	LRA4	LADH	TRC1
Zygomycota		Zygomycetes	*Rhizopus oryzae*			0	0	0	0	0	0
Basidiomycota		Ustilaginomycetes	*Ustilago maydis*	+		1	0	0	0	1	0
Basidiomycota		Hymenomycetes	*Cryptococcus neoformans*	+	C	1	0	1	0	0	1
Basidiomycota		Hymenomycetes	*Phanerochaete chrysosporium*			1	0	0	1	0	1
Basidiomycota		Hymenomycetes	*Coprinus cinereus*			0	0	0	0	0	0
Ascomycota		Schizosaccharomycetes	*Schizosaccharomyces pombe*	−		0	0	0	0	1	0
Ascomycota	Pezizomycotina	Eurotiomycetes	*Coccidioides immitis*			0	0	0	0	1	0
Ascomycota	Pezizomycotina	Eurotiomycetes	*Aspergillus nidulans*			2	1	1	0	1	1
Ascomycota	Pezizomycotina	Eurotiomycetes	*Aspergillus fumigatus*		C	1	1	1	0	2	1
Ascomycota	Pezizomycotina	Eurotiomycetes	*Aspergillus niger*		C	2	1	1	0	1	1
Ascomycota	Pezizomycotina	Eurotiomycetes	*Aspergillus oryzae*		C	1	0	1	0	1	1
Ascomycota	Pezizomycotina	Dothideomycetes	*Phaeosphaeria nodorum*			1	1	1	1	0	0
Ascomycota	Pezizomycotina	Sordariomycetes	*Magnaporthe grisea*		C	1	1	1	1	1	1
Ascomycota	Pezizomycotina	Sordariomycetes	*Neurospora crassa*		C	1	1	1	1	1	1
Ascomycota	Pezizomycotina	Sordariomycetes	*Chaetomium globosum*		C	1	1	1	1	1	1
Ascomycota	Pezizomycotina	Sordariomycetes	*Trichoderma reesei*		C	1	1	1	1	2	2
Ascomycota	Pezizomycotina	Sordariomycetes	*Fusarium graminearum*		C	1	1	1	1	1	2
Ascomycota	Pezizomycotina	Sordariomycetes	*Nectria haematococca*		C	2	1	1	1	3	2
Ascomycota	Pezizomycotina	Leotiomycetes	*Botrytis cinerea*		C	1	1	1	1	1	1
Ascomycota	Pezizomycotina	Leotiomycetes	*Sclerotinia sclerotiorum*		C	1	1	1	2	0	1
Ascomycota	Saccharomycotina	Saccharomycetes	*Yarrowia lipolytica* (*W29/CBS 7504*)	−		0	0	0	1	2	0
Ascomycota	Saccharomycotina	Saccharomycetes	*Candida guilliermondii* (*ATCC 6260/CBS 566*)	+	F	1	0	1	3	1	1
Ascomycota	Saccharomycotina	Saccharomycetes	*Candida lusitaniae* (*ATCC 42720*)	+	F	1	1	1	1	1	1
Ascomycota	Saccharomycotina	Saccharomycetes	*Debaryomyces hansenii* (*CBS767*)	w	F	1	1	1	3	2	2
Ascomycota	Saccharomycotina	Saccharomycetes	*Scheffersomyces stipitis* (*CBS 6054*)	+	F	1	1	1	3	1	1
Ascomycota	Saccharomycotina	Saccharomycetes	*Candida albicans* (*SC5314/CBS 8758*)	−		0	0	0	0	1	0
Ascomycota	Saccharomycotina	Saccharomycetes	*Saccharomyces castellii* (*NRRL Y-12630/CBS4309*)	−		0	0	0	0	0	0
Ascomycota	Saccharomycotina	Saccharomycetes	*Candida glabrata* (*CBS 138*)	−		0	0	0	0	3	0
Ascomycota	Saccharomycotina	Saccharomycetes	*Saccharomyces cerevisiae*	−		0	0	0	0	3	0
Ascomycota	Saccharomycotina	Saccharomycetes	*Saccharomyces kluyveri* (*CBS 3082*)	−		0	0	0	0	0	0
Ascomycota	Saccharomycotina	Saccharomycetes	*Ashbya gossypii*			0	0	0	0	2	0
Ascomycota	Saccharomycotina	Saccharomycetes	*Kluyveromyces lactis* (*NRRL Y-1140/CBS 2359*)	−		0	0	0	0	3	0

or only LADH. The Saccharomycotina *Candida guilliermondii*, a close relative of *C. lusitaniae*, *D. hansenii* and *S. stipitis*, lacks an *LRA2*.

2.2. Clustering of L-rhamnose catabolism genes

L-rhamnose catabolism genes, RHA1, LRA2, LRA3 and LRA4 have been previously found in a chromosomal gene cluster in *S. stipitis* (Watanabe et al., 2008b). In order to study when such cluster was formed and how it is distributed in Fungi, we searched for chromosomal regions containing the protein domains of L-rhamnose catabolism genes in 32 fungi. The full L-rhamnose catabolism gene cluster was only found from species closely related to *S. stipitis* (Fig. 2, column "Cluster" in Table 1). A putative transcription factor, *TRC1* (ABN68604), with the domains "Fungal transcriptional regulatory protein, N-terminal" (Zn(II)2Cys6 i.e. IPR001138) and "Fungal specific transcription factor" -domain (IPR007219) was found beside the *S. stipitis* L-rhamnose cluster. Orthologues of this gene were found close to *RHA1* and *LRA3* in all Pezizomycotina studied, excluding *Aspergillus nidulans* and *Phaeosphaeria nodorum*, in the basidiomycete *C. neoformans* and in Saccharomycotina species closely related to *S. stipitis*. Hence, the physical co-occurrence of only *RHA1*, *LRA3* and *TRC1* was defined as the core cluster ('C') to differentiate it from the full L-rhamnose catabolism cluster ('F', in column "Cluster" in Table 1). Interestingly, two yeasts taxonomically close to *S. stipitis*, *Spathaspora passalidarum* and *Pachysolen tannophilus*, do not appear to possess any orthologues to the L-rhamnose cluster (data not shown). In *A. nidulans* none of the L-rhamnose catabolism cluster genes are found as clustered and in *P. nodorum LRA2* and *LRA3* are side-by-side. In closely related species, such as the Eurotiomycetes *Aspergillus oryzae*, *Aspergillus niger* and *Aspergillus fumigatus* and the Sordariomycetes *Nectria haemaetococca*, *Fusarium graminearum* and *T. reesei*, the *RHA1*, *LRA3* and *TRC1* cluster (core cluster) is probably found in a syntenic region. However, between distant species, such as the basidiomycete *C. neoformans* and Ascomycota species, there is no evidence for synteny around the core cluster (Fig. 2, Supplementary Fig. 7).

Phylogenetic trees of the rhamnose cluster genes including the transcription factor were calculated for each gene separately (Supplementary Figs. 1–6). They resemble the expected species tree (Fitzpatrick et al., 2006).

2.3. L-rhamnonate dehydratase, LRA3

The *RHA1*, *LRA2*, *LRA4* and *LADH* genes have been previously characterised in *S. stipitis*. In order to finalise the biochemical characterisation of the L-rhamnose cluster, the L-rhamnonate dehydratase from *S. stipitis* was heterologously expressed in *S. cerevisiae*. For that purpose, the open reading frame was cloned in an *S. cerevisiae* expression vector with a strong constitutive promoter. In *S. stipitis*, the codon CUG is translated to serine and not to leucine as it does in *S. cerevisiae* and many other organisms (Laplaza et al., 2006). The L-rhamnonate dehydratase contained two such codons at bp 4–6 and 1120–1122. These were changed to UCG so that the heterologous protein contained the same amino acid sequence. The expression of the L-rhamnonate dehydratase in *S. cerevisiae* resulted in L-rhamnonate dehydratase activity in the crude cell extract. At an L-rhamnonate concentration of 10 mM the activity was 0.032 ± 0.004 nkat per mg of extracted protein. With L-mannonate a similar activity was observed. Also with L-lyxonate activity was observed however this activity was reduced by about 95%. No activity was observed with any of the other following sugars: L-gulonate, D-galactonate, L-galactonate,

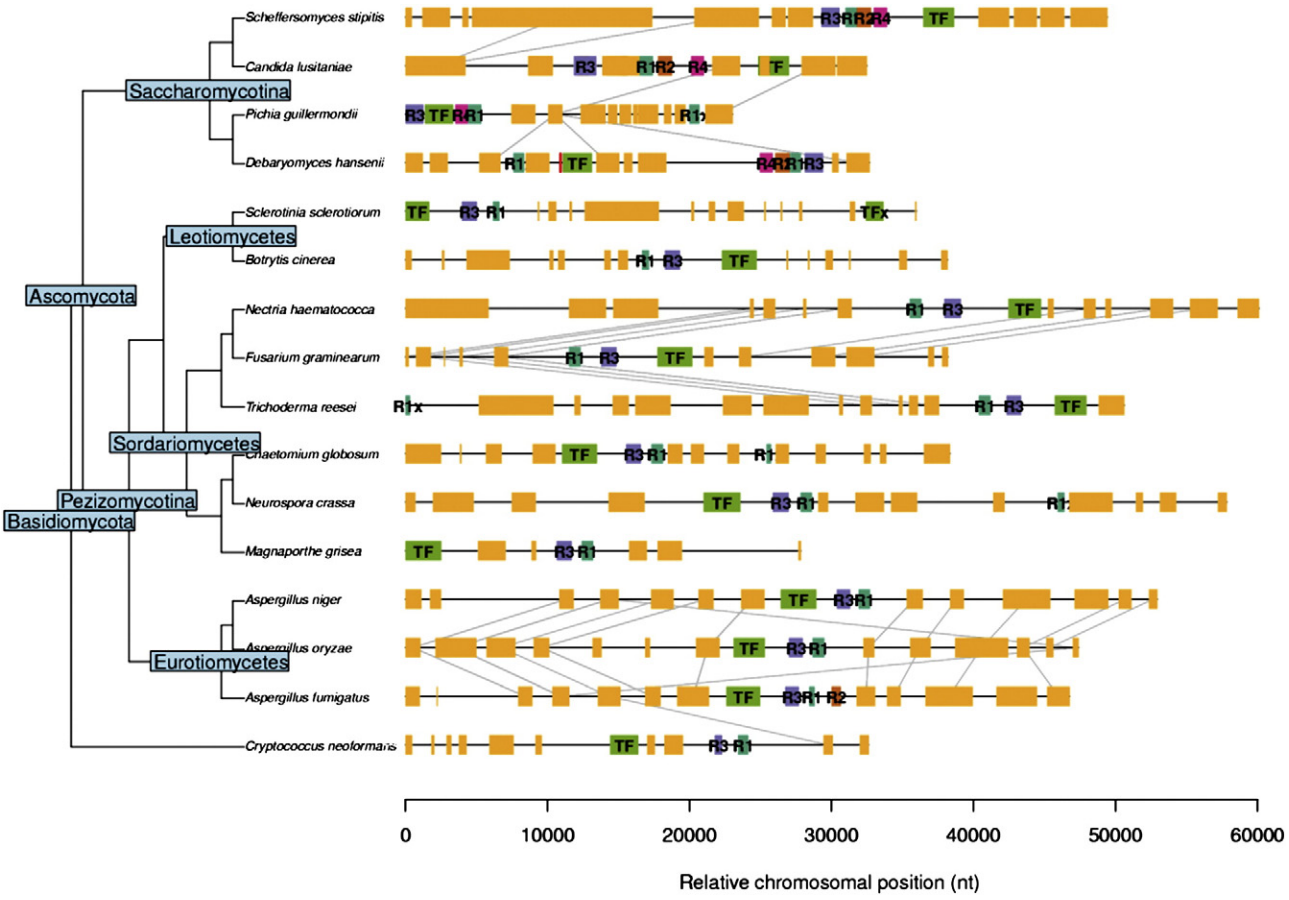

Fig. 2. L-Rhamnose gene clusters on fungal chromosomes. On the left an approximate phylogenetic tree of the species compiled from literature (Fitzpatrick et al., 2006; Jeffries et al., 2007). On the right a stretch of a scaffold from each species containing the cluster and neighbouring genes. Genes are shown as boxes on the scaffold stretch. Orthologues of RHA1 (R1), LRA2 (R2), LRA3 (R3), LRA4 (R4) and TRC1 (TF) are indicated when present and genes belonging to same families as any of them are indicated by an additional 'x'. Grey lines connect genes with identical protein domains on adjacent scaffolds (excluding R1, R2, R3, R4 and TF) in order to reveal syntenies. Additional detail in Supplementary Fig. 7.

D-fuconate, D-ribonate, L-arabonate, D-xylonate, meso-galactarate. The crude cell extract obtained from the control strain showed no activity towards any of the sugar acids. To facilitate the purification we attached a histidine-tag to the N-terminus or to the C-terminus of the protein. In order to test whether this modification of the protein had an effect on the activity we expressed the tagged protein with the same plasmid in the same *S. cerevisiae* strain and compared the activities in the crude cell extract. The C-terminally his-tagged protein showed no activity and the activity of the N-terminally tagged protein was reduced by about 95%. Since the tagging of the protein had such a severe effect on the activity in the extract we did not purify the tagged protein but instead used the crude extract to estimate the Michaelis–Menten constants. We estimated for L-rhamnonate a Km of 25 mM and a vmax 0.11 nkat per mg of protein in the crude cell extract (Fig. 3).

2.4. Aldolases

We found that the Saccharomycotina *S. stipitis*, *D. hansenii*, *C. lusitaniae* and the Pezizomycotina *Sclerotinia sclerotiorum* had additional non-clustered paralogues of *LRA4*. In order to assess the physiological role of these enzymes we compared the kinetic properties of the closest homologue of the *S. stipitis LRA4* that we called *LRA41* (ABN65879). Both proteins were heterologously expressed in *S. cerevisiae* with an N-terminal histidine-tag. The *LRA41*, had two CTG codons, which were changed to TCG before the expression in *S. cerevisiae*; the *LRA4* did not contain CTG's.

To assay the aldolase activity we assayed the reverse reaction since it is known that the enzyme is reversible and the equilibrium on the side of the 2-keto-3-deoxy-L-rhamnonate (Rigo et al., 1985). We assayed the appearance of the 2-keto-3-deoxy compound using the TBA (thiobarbituric acid) assay and the disappearance of pyruvic acid from the reaction medium. The measurement of the disappearance of pyruvic acid is required for aldehydes with no oxygen bound to the C2 since the TBA assay would not work in these cases.

Both enzymes exhibited activity with pyruvate and lactaldehyde but showed also activity with pyruvate and other aldehydes. These activities are summarised in Table 2. Both enzymes have high activity with L-lactaldehyde and lower with D-lactaldehyde. The highest

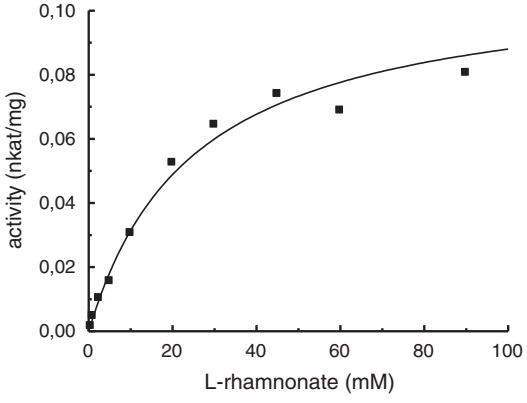

Fig. 3. Kinetic properties of the L-rhamnonate dehydratase: The heterogeneously expressed protein was analysed in a crude cell extract. The curve was calculated assuming a Michaelis–Menten kinetic model with Km = 25 mM and vmax = 0.11 nkat/mg.

activity of the *LRA41* however is with L-glyceraldehyde. Both enzymes are rather unspecific. They show activity with glycoaldehyde and glyoxal and the *LRA4* a small activity with glyoxalic acid indicating that the oxidation state of the C2 is not of great importance. Acetaldehyde and formaldehyde show activity but with a longer aliphatic chain as in butyraldehyde no activity is observed. Also aldose (hexose and pentose) sugars did not exhibit activity.

2.5. Transcriptome

It has been suggested that metabolic gene clustering arises due to benefits of co-regulation of the cluster's genes (Hurst et al., 2004). In order to verify if the L-rhamnose catabolism genes are co-regulated and asses the physiological role of LRA4 paralogues, *S. stipitis* was grown on rhamnose and analysed with oligonucleotide microarrays. Data was submitted to GEO (http://www.ncbi.nlm.nih.gov/geo/) with accession number GSE28557. Data from this experiment was then normalised together with data from Jeffries and Van Vleet (2009) in order to compare growth on L-rhamnose to growth on other carbon sources. All experiments were done with triplicate biological repeats. Genes most induced by L-rhamnose were retrieved by comparing each gene's expression signal on L-rhamnose to signal on D-glucose and filtering for genes with at least 2 fold up regulation and a p-value of $p<0.05$ for the said difference. In total 22 such genes were retrieved (Fig. 4). Pearson correlation of these 22 genes' expression on L-rhamnose in comparison to the different culture conditions from (Jeffries and Van Vleet, 2009) was calculated and significant correlation was detected only for arabinose ($p<0.0015$). Besides L-rhamnose cluster enzymes, *LADH* and *TRC1*, one putative transcription factor (*FST14*) of the IPR007219 family, 10 metabolic enzymes (*ICL1*, *PCK1*, *FBP1*, *CAT2*, *YAT1*, *ALD6*, *FAA24*, *BLG5*, *HIB1*, and *GLR2*) and three transporters (*HXT4*, *FUC1* and *SFC1*) were detected. Of the cluster's genes, *LRA4*, *LRA3*, and *RHA1* are over 10 fold induced, as is the non-clustered *LADH*. In addition, gene 33673, physically situated beside *LRA3* on chromosome 8, is induced 5 fold (Fig. 4). It is an orphan gene, for which no conserved domains, nor blast hits with E-value below 0.01, were found at the NCBI (http://blast.ncbi.nlm.nih.gov/Blast.cgi) sequence database. No other induced genes were found physically adjacent to each other.

3. Discussion

The L-rhamnose cluster in *S. stipitis* consists of the genes *RHA1* (Koivistoinen et al., 2008), *LRA2*, *LRA3* and *LRA4* (Watanabe et al., 2008b). Based on evidence of gene presence and chromosomal positioning (Table 1 and Fig. 2) it is likely that the full rhamnose usage cluster was formed in the common ancestor of *S. stipitis*, *D. hansenii*, *C. lusitaniae* and *C. guilliermondii*. The core cluster of *RHA1*, *LRA3* and *TRC1*, was likely to be already present in the common ancestor of Pezizomycotina and Saccharomycotina as it would be unlikely to form so many times independently.

Horizontal gene transfer HGT has been established as a rare, relative to prokaryotes, but important force of genome evolution in Fungi (for review (Richards et al., 2011; Richards, 2011)). In particular, HGT of metabolic clusters have been shown (Slot and Hibbett, 2007; Slot and Rokas, 2010) (Patron et al., 2007; Khaldi et al., 2008) (Slot and Rokas, 2011). Key evidence for HGT are phylogenetic trees of genes that contradict the expected phylogenetic tree of studied species. As the phylogenetic trees constructed of the rhamnose cluster genes (Supplementary Figs. 1–6) resemble the expected species tree, there is no cause to suspect HGT as the origin of *C. neoformans* core cluster. Thus, the core cluster's presence in *C. neoformans* either reflects earlier ancestry or independent emergence of the cluster. In particular, in *C. neoformans* the independent emergence of the galactose cluster has been show (Slot and Rokas, 2010).

Besides the nitrate (Slot and Hibbett, 2007) and galactose cluster (Slot and Rokas, 2010), to our knowledge there is no other published case of a fungal gene cluster that would appear in both Basidiomycota and Ascomycota. The role of the core cluster's genes in other than rhamnose metabolism is not known. However, as the core cluster's distribution in fungal species is comparable to that of the nitrate and galactose clusters its' genes relevance in metabolism could also be similar and the cluster's genes could be expected to function together on a metabolic pathway.

It has been proposed that clustering of genes would also facilitate the deletion of the whole cluster (Slot and Rokas, 2010). Our data support this idea as Saccharomycotina species tend to have all genes of the cluster or none of them. If this hypothesis is correct, given that *Yarrowia lipolytica* has only *LRA4* of the cluster's genes the cluster could have been formed after separation of *Y. lipolytica* from Saccharomycotina lineage and lost in most subsequent lineages.

In Table 1 the Saccharomycetes species that contain the full L-rhamnose cluster are also growing on L-rhamnose and the Saccharomycetes species that do not have it, do not grow. *C. guilliermondii* is lacking an *LRA2* homologue coding for the L-rhamnonic acid lactone lactonase. It might be that this activity is not essential and a spontaneous hydrolysis of the lactone is sufficient for growth or that another lactonases is carrying out the reaction. Such a lactonase might be unspecific or specific but not homologous to *LRA2*. Many of the Pezizomycotina species have the core cluster and in particular all Sordariomycetes have in addition *LRA4* and *LRA2*, but elsewhere in the genome. However, we could not find information whether these species would grow on L-rhamnose as a sole carbon source. However *A. niger* and *Botrytis cinerea* are known to catabolise pectin efficiently and since L-rhamnose is a substantial part of pectin it is likely that these species catabolise it. *Ustilago maydis* is lacking almost the complete set of cluster's genes, but it has been reported to grow on L-rhamnose (Kurtzman and Fell, 1998) suggesting that it has a completely different L-rhamnose pathway or that due to phylogenetic distance its genes were missed by our homology detection.

Zn(II)2Cys6 transcription factors, like *TRC1*, are often found in fungal metabolic gene clusters regulating the genes of the cluster. It is not clear what adaptive advantage the clustering brings, but transcriptional co-regulation is likely as a dormant cluster can be induced by induction of the transcription factor (Bergmann et al., 2007) and local chromatin modifications can parallel transcriptional activation in a cluster (Roze et al., 2007). Based on our analysis of genomes and *S. stipitis* transcriptome, *TRC1* could be a regulator of the core cluster and, possibly, a regulator of the full rhamnose catabolism cluster. However, another transcription factor *FST14* was more up-regulated than *TRC1* on rhamnose. Hence, *FST14* could be a primary regulator of full rhamnose cluster, whilst *TRC1* could be the regulator of the core cluster.

Table 2
Activities of the purified enzymes LRA4 and LRA41 with pyruvate and various aldehydes.

Substrate	Specific activity[a]	
	Lra4	Lra41
DL-lactaldehyde	2.0	1.8
D-lactaldehyde	0.3	0.3
L-lactaldehyde	1.4	1.3
D-glyceraldehyde	0.7	0.5
L-glyceraldehyde	1.1	3.7
Glycoaldehyde	0.8	0.5
Glyoxal[b]	0.06	0.4
Methylglyoxal[b]	1.3	0.4
Phenylglyoxal[b]	1.1	0.2
Glyoxylic acid[b]	0.1	0
Formaldehyde[b]	0.3	0.6
Acetaldehyde[b]	0.2	0.2
Butyraldehyde[b]	0	0

[a] $\mu mol\ mg^{-1}\ min^{-1}$.
[b] Disappearance of pyruvate was followed.

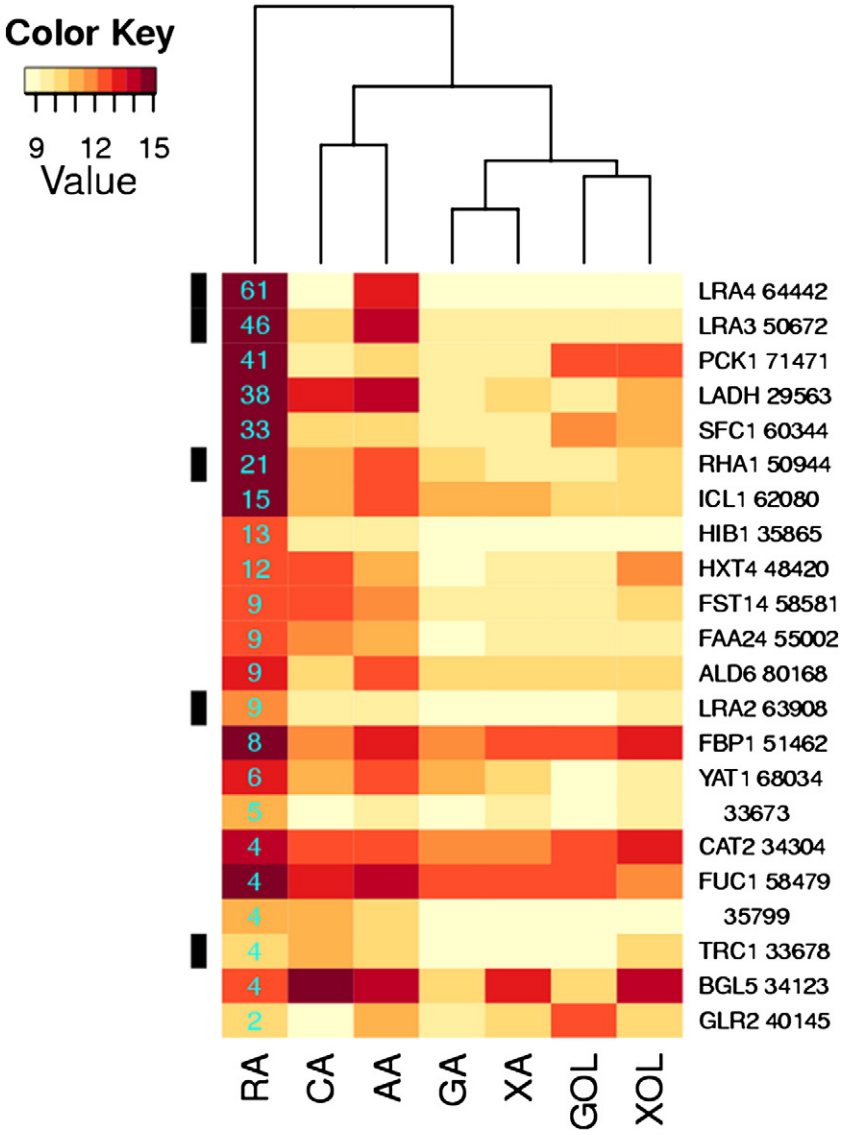

Fig. 4. Heatmap of genes with greater than 2-fold expression on RA than on GA with a p-value<0.05. Genes as rows and carbon sources as columns: GA = glucose aerobic, XA = xylose aerobic, GOL = glucose oxygen limited, XOL = XOL xylose oxygen limited, AA = arabinose aerobic, RA = rhamnose aerobic, CA = cellobiose aerobic. A colour key indicates the log2 expression signal values from individual experiments used in the cell colouring. A row side colour bar indicates the genes of the full L-rhamnose catabolism cluster. Columns are ordered by hierarchical clustering with the resulting dendrogram shown and rows by the fold change of RA relative to GA. This fold change is also shown in the respective RA cells.

In addition, it has been proposed that toxicity of intermediates drives the clustering of metabolic genes, as it facilitates both their efficient metabolism and complete removal of the pathway (Patron et al., 2007; Slot and Rokas, 2011). This theory does not contradict with the transcriptional co-regulation theory, as it also facilitates efficient metabolism and complete repression of the pathway. If a toxic metabolic intermediate would drive the clustering of L-rhamnose catabolism genes, such as intermediate would be likely to reside between RHA1 and LRA3, as they have the tightest linkage.

The activity of the enzymes of the L-rhamnose catabolism cluster had been shown previously, after they were expressed in a heterologous host (Koivistoinen et al., 2008; Watanabe et al., 2008b), except for the *LRA3*. This enzyme was expressed with an N-terminal his-tag in *E. coli* but showed no activity after purification (Watanabe et al., 2008b). We expressed it in *S. cerevisiae* where it was active, however after histidine-tagging it N-terminally or C-terminally it lost the activity or had it significantly reduced. Recently the L-rhamnonate dehydratase from *E. coli* was tagged with an N-terminal his-tag and expressed in *E. coli* (Rakus et al., 2008). Here the tagging did not have a detrimental effect. The *E. coli* enzyme had a similar specificity, i.e. it also showed activity with L-mannonic acid and with L-lyxonic acid, however the affinity for L-rhamnonic acid was much higher with the *E. coli* enzyme. The Km was about 0.15 mM compared to 25 mM for the *S. stipitis* enzyme (Rakus et al., 2008).

In the *S. stipitis* genome we identified two genes coding for enzymes with 3,6-dideoxy-L-erythro-hexulosonic acid aldolase activity, the *LRA4* and *LRA41*. The enzymes were characterised after tagging and purification. The *LRA4* is in the cluster and is up-regulated when L-rhamnose is the carbon source, whilst *LRA41* is not up-regulated, all indicating that it is indeed the *LRA4* and not the *LRA41* that is active in L-rhamnose catabolism. The activity was assayed in the reverse direction with pyruvate and aldehydes as substrate. The role of the *LRA41* remains unrevealed. The *LRA41* had highest activity with pyruvate and L-glyceraldehyde which would suggest that it could be a L-threo-3-deoxy hexulosonate aldolase, an enzyme in the fungal D-galacturonate pathway (Hilditch et al., 2007); However *S. stipitis* does not have this pathway since it is not able to catabolise D-galacturonate.

We show that *LADH* is up-regulated with L-rhamnose in agreement with earlier results (Watanabe et al., 2008a). In addition, we

find that *LADH* is never situated with the cluster and it is found in 7 species that have no other L-rhamnose catabolism genes. Thus, it is likely that its main cellular role is outside the L-rhamnose metabolism.

In conclusion we finalise the biochemical characterisation of the L-rhamnose catabolism gene cluster, provide candidate regulators based on transcriptomics and comparative genomics and a detailed analysis of its evolution in fungi. The chromosomal clustering of its core genes is a case of exceptional evolutionary conservation given our current understanding of fungal gene order conservation. Furthermore, we highlight the fact that many genes of the cluster have lineage specific paralogues. They could code for related, but distinct enzymatic activities awaiting further exploration.

4. Materials and methods

4.1. Chemicals

The DL-lactaldehyde was made according to Schoevaart (2000). D-lactaldehyde and L-lactaldehyde were prepared from L-threonine and D-threonine respectively, essentially according to the method of Huff and Rudney (1959). The resulting solution was then freeze-dried and the lactaldehyde solubilised in a small volume and applied to a Sephadex G-10 gel filtration column. The lactaldehyde concentration was estimated enzymatically using alcohol dehydrogenase from equine liver. L-glyceraldehyde was synthesised from L-gulono-1,4-lactone as described previously (Hubschwerlen, 1986; Lamble et al., 2005). The aldonic acids D-fuconate, L-rhamnonate, L-arabonate, D-xylonate and L-lyxonate were synthetised from the corresponding aldose sugars by oxidation with bromine and purified by ion exchange chromatography as described by Yew et al. (2006).

4.2. Retrieval of L-rhamnose gene cluster homologues

To find ortho- and paralogues of genes of the *S. stipitis* L-rhamnose gene cluster from fungi, their protein sequences were mapped to the in-house protein clustering database of 32 species (Arvas et al., 2007). Protein clusters that contained the genes were retrieved and sequences aligned with MAFFT version 6.857b (Katoh et al., 2002, 2005) with method L-INS-I and otherwise default options. Ortho- and paralogues for the *S. stipitis* genes were found by visually selecting the smallest homogenous region of the alignment that contained the *S. stipitis* genes. For construction of phylogenetic genes alignments were trimmed with trimAL version 1.3 with option "-automated1" (Katoh et al., 2005). Maximum likelihood (ML) phylogenetic trees were constructed with RAxML version 7.2.8 (Stamatakis, 2006). The likeliest protein substitution model was determined for each alignment with RAxML by computing the likelihood of each protein substitution model on a parsimony tree (RHA1: WAGF, LRA2: RTREVF, LRA3: RTREVF, LRA4: RTREVF, LADH: RTREVF, TRC1: JTTF). Final ML trees were constructed under the selected substitution and gamma model with 200 initial inferences and 1000 rapid bootstraps on the likeliest tree. Bootstrap support as percentage is shown on the trees (Supplementary Figs. 1–6). Trees were visualised with R (Team, 2008) library ape (Paradis et al., 2004).

4.3. Comparative analysis of L-rhamnose catabolism gene clustering

Proteins from the (Arvas et al., 2007) dataset were mapped to scaffolds by blastp. The scaffolds were then divided in windows of 12 genes that overlapped with 2 genes. Intepro protein annotations i.e. protein families, domains, active-sites etc. for each protein of a window were found from (Arvas et al., 2007) dataset. Interpro annotations of the L-rhamnose cluster genes were used to look for windows that contained these annotations. Searches were carried out and results visualised with R (Team, 2008).

4.4. L-rhamnonate dehydratase

4.4.1. Cloning

The open reading frame of the *LRA3* was cloned from genomic DNA by PCR using the following primers 1 and 2 (Table 3). In *S. stipitis* CTG codes for serine and not for leucine as it does for example in *S. cerevisiae*. In order to express in *S. cerevisiae* a gene that codes for the same amino acid sequence we changed those codons. The *LRA3* has 2 CTG's in the coding sequence at position 4–6 and 1120–1122 of the open reading frame. The codon at position 4–6 was changed to TCG in the primer that was used for cloning the open reading frame. The PCR product was then ligated to the pCR2.1 TOPO vector (Invitrogen). The resulting vector was then used to change the other CTG to TCG using the QuikChange site directed mutagenesis kit (Invitrogen). The open reading frame was released from the TOPO vector as a *BamHI* fragment and ligated to the *BamHI* site of the p2159 which is a yeast expression vector with *TPI1* promoter and *URA3* selection derived from the pYX212 (Kuorelahti et al., 2005). The histidine tagged versions of the L-rhamnonate dehydratase were made in a similar way. For the N-terminal his-tag primer 3 was used instead of primer 1 to add the amino acid sequence MHHHHHHG. For the C-terminal his-tag primer 4 was used instead of primer 2 to introduce the additional amino acid sequence GHHHHHH before the stop codon.

For the PCR the vector with the corrected codons was used as a template. The *S. cerevisiae* strain CEN.PK2-1B was then transformed with the resulting vectors and grown on selective medium lacking uracil with D-glucose as a carbon source. A control strain contained the empty expression vector p2159. To obtain cell extracts the yeast cells were lysed as described previously (Koivistoinen et al., 2008). To measure the sugar acid dehydratase activity the cell extract was incubated with the sugar acid in a buffer containing 50 mM Tris-Cl, 10 mM Mg2Cl, pH 8.0 at 30 °C. The reaction was stopped after different time intervals with trichloroacetic acid and the 2-keto-3-deoxy sugar acid quantified using the thiobarbituric acid assay as described (Buchanan et al., 1999).

4.5. Aldolases

4.5.1. Cloning

The open reading frames were cloned from genomic DNA by PCR using primers 5 and 6 for *LRA4* and primers 7 and 8 for *LRA41*. The PCR products were then ligated to the pCR2.1 TOPO vector (Invitrogen). *LRA41* has CTG codons at positions 280–282 and 583–585 which were changed to TCG using the QuikChange site directed mutagenesis kit (Invitrogen). The open reading frames were then released from the TOPO vector as a *BamHI* fragment and ligated to the *BamHI* site of the p2159 as described above. Both proteins were also tagged N-terminally with histidine-tags. The tags were introduced by PCR using the primer 9 for the *LRA4* and primer 10 for the *LRA41* to introduce the amino acid sequence MHHHHHHG before the start codon. The templates were the codon corrected TOPOvectors. The open reading frames with his-tag sequences were then released as

Table 3
Primers used.

1: GGATCCATGTCGGTAAAAGACTTCCCACAG
2: GGATCCCTAATGGCCATTCTTGATGGTTCCATTG
3: GGATCCAAAATGCATCACCATCATCACCACGGAATGTCGGTAAAAGACTTCCCAC
4: GGATCCCTAGTGGTGATGATGGTGATGTCCATGGCCATTCTTGATGGTTCC
5: GGATCCAAAATGACAATTTCAGCTGC
6: GGATCCTTATAATGTAGACTCAACACTGA
7: GGATCCAGTATGACTTTCAATAATAGCTTT
8: GGATCCTTATAAAGAGTTTTCGTATTTAAGTAAT
9: GGATCCATGCATCACCATCACCATCACGGAATGACAATTTCAGCTGCCCT
10: GGATCCATGCATCACCATCACCATCACGGAATGACTTTCAATAATAGCTTTG

BamHI fragments and ligated to the p2159. The plasmids were then transformed to *S. cerevisiae* and cell extracts prepared as described above. The activity was tested by incubating the cell extract with 2.5 mM DL-lactaldehyde and 15 mM pyruvate for different time intervals. The reaction was quenched with TCA and the formation of the 2-keto-3-deoxy sugar acid was measured with the TBA (thiobarbituric acid) assay as described in Buchanan et al. (1999). The N-terminal tag had no effect on the activity which was about 2.5 nkat/mg in the crude cell extract. The histidine tagged proteins were purified from the crude cell extract by using NiNTA resin (Qiagen). An alternative way to measure the reaction was to follow the disappearance of pyruvate. Pyruvate was measured from the TCA quenched samples enzymatically with lactate dehydrogenase. A fraction of the sample was pipetted to a medium containing 200 mM Tris–HCl pH 8.0, 200 μM NADH and lactate dehydrogenase. The pyruvate concentration was calculated from the change in NADH absorbance.

4.6. Transcriptome profiling

4.6.1. Inoculum preparation

S. stipitis CBS 6054 was cultivated aerobically for all inocula. Precultures for all cultivations were grown on YP medium (20 g/l yeast extract and 10 g/l peptone) supplemented with glucose (20 g/l). Following overnight growth, the cultures were harvested via centrifugation at 5000 rpm for 5 min in an Eppendorf 5804R centrifuge equipped with an A-4-44 rotor. Cells were rinsed once in sterile 0.9% saline and repelleted. Cell pellets were then resuspended in approximately 10 ml of sterile water. OD 600 nm measurements were taken and the correct inoculum size calculated for a final OD600 of 0.1 in the bioreactor.

4.6.2. Bioreactor cultivations

All cultivations for expression arrays were performed in New Brunswick Scientific Bioflo 110 3-L bioreactors with working volumes of 2 l each. The bioreactors were equipped with two impellers rotating at 750 rpm. Bioreactor temperature was controlled at 30 °C and the pH was kept constant at 5.0 by automatic addition of 5 N KOH. A defined minimal medium containing trace metal elements and vitamins in the following was used in all bioreactor cultivations: 2.4 g urea l^{-1}; 3 g KH2 PO4 l^{-1}, 0.5 g MgSO4 7H2O l^{-1}; 1 ml trace element solution l^{-1}; 1 ml vitamin solution l^{-1}; and 0.05 ml antifoam 289 (Sigma A-8436) l^{-1} (Modified from Verduyn et al. (Verduyn et al., 1992)). For L-rhamnose cultivations, a starting concentration of 20 g l^{-1} sugar was used. Airflow into the bioreactors for aerobic cultivations was 1 vvm. An aerobic transcriptomics sample was taken during the exponential growth phase. Cultivations and sample collections for each carbon source were performed in triplicate.

4.6.3. Transcriptomic sampling and sample processing

Transcriptomics samples were centrifuged for 5 min at 5000 rpm in an Eppendorf 5804R centrifuge equipped with an A-4-44 rotor. The supernatant solutions were removed, and the cell pellets flash frozen in liquid nitrogen. Cell pellets were stored at −80 °C prior to RNA extraction. RNA was extracted using the Qiagen RNEasy Maxi kit (Qiagen). Cell breakage was accomplished using glass beads in a BeadBeater (Biospec Products) equipped with a 15 ml chamber. RNA was later further purified and concentrated using the Qiagen RNEasy Mini kit (Qiagen). Total RNA was submitted to the University of Wisconsin, Madison Gene Expression Center for cDNA synthesis, and cDNA samples were submitted to Nimblegen for labelling and chip hybridization.

4.6.4. Expression array

The expression array used for all samples in this study was a custom array for *S. stipitis* designed by Nimblegen. This array contained 13 60-mer probes for each open reading frame (ORF) identified as of December of 2006. Each probe was replicated five times on the chip. Expressing signals from the L-rhamnose experiment and experiments of (20) were normalised together with Robust Multichip Average RMA (Irizarry et al., 2003) and significantly regulated genes (D-glucose aerobic versus L-rhamnose aerobic), were identified by Student's *T*-test (95%).

Acknowledgements

This work was financially supported by the Academy of Finland through the research programme “White Biotechnology — Green Chemistry”, The Finnish Centre of Excellence Programme, 2008–2013, Project No. 118573, an Academy Research Fellowship for P.R and an Academy Postdoctoral Researcher's fellowship 127715 of Mikko Arvas. Andreas Petrich is acknowledged for preparing the sugar acids and assaying the aldolases and Dana J. Wohlbach is gratefully acknowledged for normalising the expression array data.

Appendix A. Supplementary data

Supplementary data to this article can be found online at doi:10.1016/j.gene.2011.10.031.

References

Arvas, M., Kivioja, T., Mitchell, A., Saloheimo, M., Ussery, D., Penttilä, M., Oliver, S., 2007. Comparison of protein coding gene contents of the fungal phyla Pezizomycotina and Saccharomycotina. BMC Genomics. 8, 325.

Baldoma, L., Aguilar, J., 1988. Metabolism of L-fucose and L-rhamnose in *Escherichia coli*: aerobic–anaerobic regulation of L-lactaldehyde dissimilation. J. Bacteriol. 170, 416–421.

Bergmann, S., Schumann, J., Scherlach, K., Lange, C., Brakhage, A.A., Hertweck, C., 2007. Genomics-driven discovery of PKS-NRPS hybrid metabolites from *Aspergillus nidulans*. Nat. Chem. Biol. 3, 213–217.

Buchanan, C.L., Connaris, H., Danson, M.J., Reeve, C.D., Hough, D.W., 1999. An extremely thermostable aldolase from *Sulfolobus solfataricus* with specificity for non-phosphorylated substrates. Biochem. J. 343, 563–570.

Cooper, T.G., 1996. Regulation of allantoin catabolism in *Saccharomyces cerevisiae*. In: Marzluf, G.A. (Ed.), The Mycota III: Biochemistry and Molecular Biology. Springer, Berlin, pp. 139–169.

Covert, S.F., Vanden Wymelenberg, A., Cullen, D., 1992. Structure, organization, and transcription of a cellobiohydrolase gene cluster from Phanerochaete chrysosporium. Appl. Environ. Microbiol. 58, 2168–2175.

Fillinger, S., Felenbok, B., 1996. A newly identified gene cluster in *Aspergillus nidulans* comprises five novel genes localized in the alc region that are controlled both by the specific transactivator AlcR and the general carbon-catabolite repressor CreA. Mol. Microbiol. 20, 475–488.

Fitzpatrick, D.A., Logue, M.E., Stajich, J.E., Butler, G., 2006. A fungal phylogeny based on 42 complete genomes derived from supertree and combined gene analysis. BMC Evol. Biol. 6, 99.

Giles, N.H., Geever, R.F., Asch, D.K., Avalos, J., Case, M.E., 1991. Organization and regulation of the Qa (quinic acid) genes in *Neurospora crassa* and other fungi. J. Hered. 82, 1–7.

Gomi, K., Akeno, T., Minetoki, T., Ozeki, K., Kumagai, C., N. O, Iimura, Y., 2000. Molecular cloning and characterization of a transcriptional activator gene, amyR, involved in the amylolytic gene expression in *Aspergillus oryzae*. Biosci. Biotechnol. Biochem. 64, 816–827.

Hilditch, S., Berghäll, S., Kalkkinen, N., Penttilä, M., Richard, P., 2007. The missing link in the fungal D-galacturonate pathway: identification of the L-threo-3-deoxy-hexulosonate aldolase. J. Biol. Chem. 282, 26195–26201.

Hittinger, C.T., Rokas, A., Caroll, S.B., 2004. Parallel inactivation of multiple GAL pathway genes and ecological diversification in yeasts. PNAS 101, 14144–14149.

Hubschwerlen, C., 1986. A convenient synthesis of L-(S)-glyceraldehyde acetonide from L-ascorbic acid. Synth. Commun. 11, 962–964.

Huff, E., Rudney, H., 1959. The enzymatic oxidation of 1,2-propanediol phosphate to acetol phosphate. J. Biol. Chem. 234, 1060–1064.

Hull, E.P., Green, P.M., Arst, J., Scazzocchio, C., 1989. Cloning and physical characterization of the L-proline catabolism gene cluster of *Aspergillus nidulans*. Mol. Microbiol. 3, 553–559.

Hurst, L.D., Pál, C., Lercher, M.J., 2004. The evolutionary dynamics of eukaryotic gene order. Nat. Rev. Genet. 5, 299–310.

Irizarry, R.A., Bolstad, B.M., Collin, F., Cope, L.M., Hobbs, B., Speed, T.P., 2003. Summaries of Affymetrix GeneChip probe level data. Nucleic Acids Res. 31, e15.

Jeffries, T.W., Van Vleet, J.R., 2009. *Pichia stipitis* genomics, transcriptomics, and gene clusters. FEMS Yeast Res. 9, 793–807.

Jeffries, T.W., Grigoriev, I.V., Grimwood, J., Laplaza, J.M., Aerts, A., Salamov, A., Schmutz, J., Lindquist, E., Dehal, P., Shapiro, H., Jin, Y.S., Passoth, V., Richardson, P.M., 2007.

Genome sequence of the lignocellulose-bioconverting and xylose-fermenting yeast *Pichia stipitis*. Nat. Biotechnol. 25, 319–326.
Johnstone, I.L., McCabe, P.C., Greaves, P., Gurr, S.J., Cole, G.E., Brow, M.A., Unkles, S.E., Clutterbuck, A.J., Kinghorn, J.R., Innis, M.A., 1990. Isolation and characterisation of the *crnA-niiA-niaD* gene cluster for nitrate assimilation in *Aspergillus nidulans*. Gene 90, 181–192.
Katoh, K., Misawa, K., Kuma, K., Miyata, T., 2002. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. Nucleic Acids Res. 30, 3059–3066.
Katoh, K., Kuma, K., Toh, H., Miyata, T., 2005. MAFFT version 5: improvement in accuracy of multiple sequence alignment. Nucleic Acids Res. 33, 511–518.
Keller, N.P., Turner, G., Bennett, J.W., 2005. Fungal secondary metabolism-from biochemistry to genomics. Nat. Rev. Microbiol. 3, 937–947.
Khaldi, N., Collemare, J., Lebrun, M.H., Wolfe, K.H., 2008. Evidence for horizontal transfer of a secondary metabolite gene cluster between fungi. Genome Biol. 9, R18.
Koivistoinen, O.M., Hilditch, S., Voutilainen, S.P., Boer, H., Penttilä, M., Richard, P., 2008. Identification in the yeast *Pichia stipitis* of the first L-rhamnose-1-dehydrogenase gene. FEBS J. 275, 2482–2488.
Kuorelahti, S., Kalkkinen, N., Penttilä, M., Londesborough, J., Richard, P., 2005. Identification in the mold *Hypocrea jecorina* of the first fungal D-galacturonic acid reductase. Biochemistry 44, 11234–11240.
Kurtzman, C.P., Fell, J.W., 1998. The Yeasts: A Taxonomic Study, 4th ed. Elsevier.
Lamble, H.J., Danson, M.J., Hough, D.W., Bull, S.D., 2005. Engineering stereocontrol into an aldolase-catalysed reaction. Chem. Commun. (Camb) 124–126 Epub 2004 Nov 23.
Laplaza, J.M., Rivas Torres, B., Jin, Y.S., Jeffries, T.W., 2006. *Sh ble* and Cre adapted for functional genomics and metabolic engineering of *Pichia stipitis*. Enzyme Microb. Technol. 38, 741–747.
Lee, J.M., Sonnhammer, E.L.L., 2003. Genomic gene clustering analysis of pathways in eukaryotes. Genome Res. 13, 875–882.
Martchenko, M., Levitin, A., Hogues, H., Nantel, A., Whiteway, M., 2007. Transcriptional rewiring of fungal galactose-metabolism circuitry. Curr. Biol. 17.
Martinez, D., Berka, R.M., Henrissat, B., Saloheimo, M., Arvas, M., Baker, S.E., Chapman, J., Chertkov, O., Coutinho, P.M., Cullen, D.a.o., 2008. Genome sequencing and analysis of the biomass-degrading fungus *Trichoderma reesei* (syn. *Hypocrea jecorina*). Nat. Biotechnol. 26, 553–560.
Moralejo, P., Egan, S.M., Hidalgo, E., Aguilar, J., 1993. Sequencing and characterization of a gene cluster encoding the enzymes for L-rhamnose metabolism in *Escherichia coli*. J. Bacteriol. 175, 5585–5594.
Paradis, E., Claude, J., Strimmer, K., 2004. APE: analyses of phylogenetics and evolution in R language. Bioinformatics 20, 289–290.
Patron, N.J., Waller, R.F., Cozijnsen, A.J., Straney, D.C., Gardiner, D.M., Nierman, W.C., Howlett, B.J., 2007. Origin and distribution of epipolythiodioxopiperazine (ETP) gene clusters in filamentous ascomycetes. BMC Evol. Biol. 7, 174.
Rakus, J.F., Fedorov, A.A., Fedorov, E.V., Glasner, M.E., Hubbard, B.K., Delli, J.D., Babbitt, P.C., Almo, S.C., Gerlt, J.A., 2008. Evolution of enzymatic activities in the enolase superfamily: L-rhamnonate dehydratase. Biochemistry 47, 9944–9954.
Richards, T.A., 2011. Genome evolution: horizontal movements in the fungi. Curr. Biol. 21, 166–168.
Richards, T.A., Leonard, G., Soanes, D.M., Talbot, N.J., 2011. Gene transfer into the fungi. Fungal Biol. Rev. 25, 98–110.
Rigo, L.U., Nakano, M., Veiga, L.A., Feingold, D.S., 1976. L-Rhamnose dehydrogenase of *Pullularia pullulans*. Biochim. Biophys. Acta 445, 286–293.
Rigo, L.U., Marechal, L.R., Vieira, M.M., Veiga, L.A., 1985. Oxidative pathway for L-rhamnose degradation in *Pullularia pullulans*. Can. J. Microbiol. 31, 817–822.
Roze, L.V., Arthur, A.E., Hong, S.Y., Chanda, A., Linz, J.E., 2007. The initiation and pattern of spread of histone H4 acetylation parallel the order of transcriptional activation of genes in the aflatoxin cluster. Mol. Microbiol. 66, 713–726.
Sawada, H., Takagi, Y., 1964. The metabolism of L-rhamnose in *Escherichia coli*. 3. L-Rhamulose-phosphate aldolase. Biochim. Biophys. Acta 92, 26–32.
Schoevaart, R., 2000. Applications of Aldolases in Organic Synthesis. Technical University of Delft, Delft, The Netherlands.
Slot, J.C., Hibbett, D.S., 2007. Horizontal transfer of a nitrate assimilation gene cluster and ecological transitions in fungi: a phylogenetic study. PLoS One 2, e1097.
Slot, J.C., Rokas, A., 2010. Multiple GAL pathway gene clusters evolved independently and by different mechanisms in fungi. Proc. Natl. Acad. Sci. U. S. A. 107, 10136–10141.
Slot, J.C., Rokas, A., 2011. Horizontal transfer of a large and highly toxic secondary metabolic gene cluster between fungi. Curr. Biol. 21, 134–139.
Stamatakis, A., 2006. RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. Bioinformatics 22, 2688–2690.
Takagi, Y., Sawada, H., 1964a. The metabolism of L-rhamnose in *Escherichia coli*. I. L-Rhamnose isomerase. Biochim. Biophys. Acta 92, 10–17.
Takagi, Y., Sawada, H., 1964b. The metabolism of L-rhamnose in *Escherichia coli*. II. L-Rhamnulose kinase. Biochim. Biophys. Acta 92, 18–25.
Team, R.D.C., 2008. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria.
Twerdochlib, A.L., Pedrosa, F.O., Funayama, S., Rigo, L.U., 1994. L-Rhamnose metabolism in *Pichia stipitis* and *Debaryomyces polymorphus*. Can. J. Microbiol. 40, 896–902.
Verduyn, C., Postma, E., Scheffers, W.A., Van Dijken, J.P., 1992. Effect of benzoic acid on metabolic fluxes in yeasts: a continuous-culture study on the regulation of respiration and alcoholic fermentation. Yeast 8, 501–517.
Vieira, M.M., Rigo, L.U., Marechal, L.R., Veiga, L.A., 1979. Induction and catabolite repression of L-rhamnose dehydrogenase in *Pullularia pullulans*. J. Bacteriol. 138, 55–59.
Watanabe, S., Piyanart, S., Makino, K., 2008a. Metabolic fate of L-lactaldehyde derived from an alternative L-rhamnose pathway. FEBS J. 11, 11.
Watanabe, S., Saimura, M., Makino, K., 2008b. Eukaryotic and bacterial gene clusters related to an alternative pathway of nonphosphorylated L-rhamnose metabolism. J. Biol. Chem. 283, 20372–20382.
Wilson, D.M., Ajl, S., 1957a. Metabolism of L-rhamnose by *Escherichia coli*. I. L-Rhamnose isomerase. J. Bacteriol. 73, 410–414.
Wilson, D.M., Ajl, S., 1957b. Metabolism of L-rhamnose by *Escherichia coli*. II. The phosphorylation of L-rhamnulose. J. Bacteriol. 73, 415–420.
Yew, W.S., Fedorov, A.A., Fedorov, E.V., Rakus, J.F., Pierce, R.W., Almo, S.C., Gerlt, J.A., 2006. Evolution of enzymatic activities in the enolase superfamily: L-fuconate dehydratase from *Xanthomonas campestris*. Biochemistry 45, 14582–14597.
Yuan, X.L., Roubos, J.A., C.A.M.J.J., V.D.H., Ram, A.F.J., 2008. Identification of InuR, a new Zn(II)2Cys6 transcriptional activator involved in the regulation of inulinolytic genes in *Aspergillus niger*. Mol. Genet. Genet. 279, 11–26.