Comparison of the paralogous transcription factors AraR and XlnR in *Aspergillus oryzae*

Kana Ishikawa [a], Emi Kunitake [a,b], Tomomi Kawase [a], Motoki Atsumi [a], Yuji Noguchi [a], Shuhei Ishikawa [a], Masahiro Ogawa [c], Yasuji Koyama [c], Makoto Kimura [a], Kyoko Kanamaru [a], Masashi Kato [a,d], and Tetsuo Kobayashi [a, #]

[a] Department of Biological Mechanisms and Functions, Graduate School of Bioagricultural Sciences, Nagoya University, Furo-cho, Chikusa-ku, Nagoya-shi, Aichi 464-8601, Japan; [b] Department of Life Sciences, Graduate School of Bioresources, Mie University, 1577 Kurimamachiya‐cho, Tsu-shi, Mie 514‐8507, Japan; [c] Noda Institute for Scientific Research, 399 Noda, Noda-shi, Chiba 278-0037, Japan; [d] Faculty of Agriculture, Meijo University, Tenpaku-ku, Nagoya-shi, Aichi 468-8502, Japan

[#] Address correspondence to Tetsuo Kobayashi

Phone: +81-52-789-4085

Fax: +81-52-789-4087

E-mail: koba@agr.nagoya-u.ac.jp

ORCID: 0000-0002-4008-454X

Abstract

The paralogous transcription factors AraR and XlnR in *Aspergillus* regulate genes that are involved in degradation of cellulose and hemicellulose and catabolism of pentose. AraR and XlnR target the same genes for pentose catabolism but target different genes encoding enzymes for polysaccharide degradation. To uncover the relationship between these paralogous transcription factors, we examined their contribution to regulation of the PCP genes and compared their preferred recognition sequences. Both AraR and XlnR are involved in induction of all the pentose catabolic genes in *A. oryzae* except *larA* encoding L-arabinose reductase, which was regulated by AraR but not by XlnR. DNA binding studies revealed that the recognition sequences of AraR and XlnR also differ only slightly; AraR prefers CGGDTAAW, while XlnR prefers CGGNTAAW. All the pentose catabolic genes possess at least one recognition site to which both AraR and XlnR can bind. Cooperative binding by the factors was not observed. Instead, they competed to bind to the shared sites. XlnR bound to the recognition sites mentioned above as a monomer, but bound to the sequence TTAGSCTAA on the xylanase promoters as a dimer. Consequently, AraR and XlnR have significantly similar, but not the same, DNA-binding properties. Such a slight difference in these paralogous transcription factors may lead to complex outputs in enzyme production depending on the concentrations of co-existing inducer molecules in the natural environment.

Keywords: *Aspergillus oryzae*, pentose catabolism, AraR, XlnR

Introduction

Filamentous fungi produce a variety of enzymes to degrade lignocellulose, such as xylanases, mannanases, and cellulases. Expression of the genes encoding those enzymes is regulated by various transcription factors, and a specific set of the genes is upregulated in response to environmental conditions (Benocci et al. 2017; Gruben et al. 2017; Huberman et al. 2016; Kunitake and Kobayashi 2017; Tani et al. 2014).

XlnR is a Zn_2Cys_6 transcription factor that is essential to production of xylanolytic enzymes in *Aspergillus* (Marui et al. 2002b; van Peij et al. 1998b) and contributes to regulation of genes that are involved in the pentose catabolic pathway (PCP) (Battaglia et al. 2011a, b;Battaglia et al. 2014; Hasper et al. 2000; Noguchi et al. 2009). Although their physiological roles differ across organisms (Klaubauf et al. 2014), orthologs of XlnR are widely conserved among ascomycetous fungi, including species in the genera *Aspergillus*, *Penicillium*, *Neurospora*, *Fusarium*, and *Trichoderma* (Benocci et al. 2017; Huberman et al. 2016; Tani et al. 2014). Common targets of XlnR orthologs are genes encoding xylanolytic enzymes and genes involved in the PCP, except in *Magnaporthe oryzae*, wherein XlnR orthologs regulate PCP genes but not genes encoding xylanolytic enzymes (Battaglia et al. 2013). In addition to XlnR, filamentous fungi in the class *Eurotiomycetes* generally have its paralog AraR (Benocci et al. 2017; Huberman et al. 2016; Tani et al. 2014). In *Aspergillus*, AraR regulates the PCP genes that XlnR regulates as well as genes that are not regulated by XlnR (Battaglia et al. 2011a, b; Battaglia et al. 2014; de Souza et al. 2013; de Vries et al. 2011; Kowalczyk et al. 2015). Thus, there are AraR-specific, XlnR-specific, and AraR/XlnR-regulated genes. However, regulation by AraR and XlnR does not appear to be so simple. For example, both D-xylose and L-arabinose induce both AraR- and XlnR-dependent expression of the *xyrA*, *xdhA*, and *xkiA* genes in *A. nidulans*. But XlnR also represses their expression in response to L-arabinose, as suggested by the observations that the expression of those genes is higher in a Δ*xlnR* mutant than

they are in a reference strain (Kowalczyk et al. 2015) and that *larA*, *lxrA*, and *xkiA* are not induced in *A. niger* in the presence of both D-xylose and L-arabinose (de Souza et al. 2013). These observations suggest that XlnR and AraR mutually interfere with one another via an unknown mechanism.

Zn_2Cys_6 transcription factors are specific to fungi and generally bind as dimers to an inverted, everted, or direct repeat of a CGG triplet (Todd and Andrianopoulos 1997). Exceptionally, XlnR is reported to bind to the sequence GGCTAAA and related sequences (de Vries et al. 2002; Marui et al. 2002b; van Peij et al. 1998a, b). The recognition sequence of AraR has not yet been identified experimentally, although Battaglia et al. (2014) report that the promoters of AraR-regulated PCP genes in *A. niger* contain six conserved motifs, one of which has a sequence that is highly similar to the consensus sequence in the recognition sequence of XlnR.

In this manuscript, we report a study in which we compared AraR and XlnR with particular focus on their recognition sequences. Our results indicate that their recognition sequences are very similar, and they competed for binding to shared binding sites. Slight differences in the otherwise overlapping functions of AraR and XlnR may lead to differential expression of their target genes depending on the concentrations of their inducers coexisting in the natural environment.

Materials and Methods

Strains used in this study

A. oryzae strain RIB40 was used as the wild type control in this study. *A. oryzae* RkuN16ptr1, which is *pyrG* and *ku70* knockout of RIB40 (Takahashi et al. 2006a, b), was used as the parental strain for constructing Δ*araR*, Δ*xlnR*, and Δ*araR*/Δx*lnR* mutants. *E. coli* XL1-Blue was used for gene manipulations, and *E. coli* BL21(DE3) and BL21(DE3)-pLysS were used to produce recombinant proteins.

Production and purification of *A. oryzae* pentose catabolic enzymes in *E. coli*

Enzymes in the pentose catabolic pathway were expressed as His_6-tagged recombinant proteins in *E. coli*. The *xyrA*, *ladA*, and *xdhA* genes encoding D-xylose reductase, L-arabinitol 4-dehydrogenase, and xylitol dehydrogenase, respectively, of *A. oryzae* were identified previously (Kaneda et al. 2011; Suzuki et al. 2005; Tran et al. 2004). The candidate genes for *larA* (AO090023000264 in AspGD data base, http://www.aspergillusgenome.org/) encoding L-arabinose reductase and *lxrA* (AO090038000426) encoding L-xylulose reductase were selected based on DNA microarray analysis (de Vries et al. 2011) and sequence similarity to *larA* and *lxrA* from *A. niger* (Mojzita et al. 2010a, b). cDNA encoding the enzymes were obtained by RT-PCR from total RNA from *A. oryzae* RIB40. Primers used for PCR amplifications are listed in Table S1. The amplified fragments were digested with restriction enzymes at sites introduced on the PCR primers, ligated with the pET-15b vector (Merck Millipore Co., Darmstadt, Germany) digested with the same enzymes, and then transformed into *E. coli* XL1-Blue. Then, the resulting recombinant plasmids were introduced into *E. coli* BL21(DE3)-pLysS for expression of the enzymes as His_6-tagged proteins.

Expression of recombinant proteins in *E. coli* BL21 (DE3) pLysS was carried out following instructions provided by the manufacturer of the pET system. The recombinant proteins were purified by affinity column chromatography using Ni-NTA agarose (Qiagen, Hilden, Germany). The recombinant proteins were eluted with a Tris-HCl, pH 8.0, buffer containing 50 mM NaCl, 1 mM EDTA, and 100 mM imidazole, and then dialyzed against 50 mM Tris-HCl, pH 8.0 (Fig. S1). Concentrations of purified recombinant proteins were determined by the Lowry method (Lowry et al. 1951).

Enzyme assay

The enzyme activities of the recombinant proteins were measured as described previously (de Groot et al. 2005). Reaction mixtures consisted of 0.4 mM $NAD(P)^+$ and 50 mM substrate in a 100 mM glycine/NaOH buffer, pH 9.6, for pentitol oxidation, and 0.2 mM NAD(P)H and 50 mM substrate in a 50 mM Na-phosphate buffer, pH 7.6, for pentose reduction. Enzyme reactions were carried out at 30°C and changes in absorbance at 340 nm were monitored with a V-560 UV/visible spectrophotometer (JASCO Co., Tokyo, Japan). One unit of enzyme activity was defined as the amount of enzyme needed to reduce 1 μmol of $NAD(P)^+$ or oxidize 1 μmol of NAD(P)H per min.

Optimal pH values for the enzymes were determined using McIlvaine buffer (pH 4.5 – 8.0) and 100 mM glycine/NaOH buffer (pH 8.0 – 11.0). K_m and V_{max} values were determined using the Lineweaver-Burk plot.

Deletion of *araR* and/or *xlnR*

The Δ*araR* and Δ*xlnR* mutants were constructed by replacing the target gene with the *A. oryzae pyrG* gene via homologous recombination. DNA fragments for the deletions were produced by overlap extension PCR as described previously (Ogawa et al. 2010). Chromosomal DNA from *A. oryzae* RIB40 was used as the template for amplification of the upstream and downstream regions of the *araR* and *xlnR* genes. Plasmid pYRB9, which carries a 3-kb *Bgl*II fragment containing *pyrG*, was used as the template for amplification of the *pyrG* gene. Primers used for PCR amplifications are listed in Table S2. The Δ*araR*Δ*xlnR* double deletion strain was constructed using the marker recycling strategy (Akada et al. 2006). First, the *xlnR* gene was deleted with *pyrG* as the marker using a cassette produced by fusion PCR with the primers designed for marker recycling, shown in Table S2. Then, after the Δ*xlnR* mutant lacking the *pyrG* maker gene, which shows 5-fluoroorotic acid resistance, was obtained, *araR* was also deleted from it as described above.

Transcriptional analysis

The effects of *araR* and/or *xlnR* deletions performed in this study on expression of pentose catabolic genes were determined by RT-qPCR analysis. Strains RIB40 and RkuN16ptr1 and the Δ*xlnR*, Δ*araR*, and Δ*araR*/Δ*xlnR* mutants were grown for 16 h at 30°C in Czapek Dox medium in which carbon and nitrogen sources were replaced with 1% Hipolypepton (Nihon Pharmaceutical, Tokyo, Japan) and 0.5% Bacto™ Yeast Extract (BD, New Jersey, USA). The mycelia were harvested and washed with Czapek Dox medium lacking a carbon source and then transferred to Czapek Dox medium containing either 5.6 mM fructose (un-inducing conditions) or 5.6 mM fructose with 25 mM D-xylose or L-arabinose (inducing conditions). After incubating for 1 h, the mycelia were harvested and ground into fine powder in liquid nitrogen. RNA extraction and RT-qPCR were performed as described previously (Kunitake et al. 2016) using the primers shown in Table S3.

DNA binding studies of AraR and XlnR

The N-terminal regions of AraR (aa 1-84) and XlnR (aa 1-183) that carry the $Zn(II)_2Cys_6$ DNA-binding motif were expressed as His_6-tagged (AraR84-H) and FLAG-tagged (F-XlnR183) proteins, respectively, in *E. coli*. Briefly, DNA fragments encoding the regions were PCR-amplified. A cDNA mixture from *A. oryzae* RIB40 was used as the template for amplification of *araR* and plasmid pAF4 (Noguchi et al. 2011) was used as the template for amplification of *xlnR*. The primer sets used are shown in Table S4. The amplified DNA fragment encoding *araR* was digested with *Nco* I and *Sac* I and that encoding *xlnR* was digested with *Hin*d III and *Bgl* II and then inserted into the corresponding sites on pET52b(+) (Merck Millipore Co., Darmstadt, Germany) and pT7-FLAG™-1 (Sigma-Aldrich Co. LLC., St. Louis, MO, USA), respectively. The resulting recombinant plasmids were introduced into *E. coli* BL21(DE3).

Recombinant proteins were purified by affinity column chromatography using Ni-NTA agarose

(Qiagen, Hilden, Germany) or ANTI-FLAG M2 Antibody Affinity Gel (Sigma-Aldrich Co. LLC., St. Louis, MO, USA) according to manufacturer's instructions. Solution of purified proteins were dialyzed against a buffer consisting of 200 mM NaCl, 1 mM EDTA, 50% glycerol, and 20 mM Tris-HCl, pH 7.4 (Fig. S1).

DNA binding studies of the recombinant AraR and XlnR proteins were carried out by electrophoretic mobility shift assays (EMSA) as described previously (Li et al. 2016). DNA fragments corresponding to the promoter regions of the PCP genes were amplified with primer sets listed in Table S5 and then digested with the restriction enzymes at sites added on the primers. DNA fragments were labeled by fill-in reaction using Klenow fragment and dNTP mixture containing biotin-14-dCTP (Thermo Fisher Scientific Inc., Waltham, MA, USA). To mutagenize candidate binding sites, the amplified fragments were first subcloned into pBluescript II KS(+). Mutations were then introduced by QuikChange Site-Directed Mutagenesis (Agilent Technologies Santa Clara, CA, USA).

Short probes (less than 100 bp) were prepared by annealing the oligonucleotide pairs listed in Table S6 and then labeled with biotin-14-dCTP by fill-in reaction as described above. EMSA was carried out using the LightShift^{TM} EMSA Optimization and Control Kit (Thermo Fisher Scientific Inc., Waltham, MA, USA) according to the manufacturer's protocol. A LAS-3000 Mini imaging system (Fuji Photo Film Co., Ltd., Tokyo, Japan) was used for signal detection.

Results

Confirmation of pentose catabolic pathway (PCP) genes in *A. oryzae*

In the pentose catabolic pathway in *Aspergillus*, conversion of D-xylose to D-xylulose is catalyzed by D-xylose reductase (XyrA) and xylitol dehydrogenase (XdhA), while L-arabinose is converted to D-xylulose by sequential enzyme reactions mediated by L-arabinose reductase (LarA),

L-arabinitol 4-dehydrogenase (LadA), L-xylulose reductase (LxrA), and finally XdhA. Then, D-xylulose is phosphorylated by D-xylulokinase A (XkiA) and enters the pentose phosphate pathway (Mojzita et al. 2010a, b). The *xyrA* (AO090003000859 in the AspGD database, http://www.aspergillusgenome.org/), *ladA* (AO090005001078), and *xdhA* (AO090038000631) genes in *A. oryzae* have been identified (Kaneda et al. 2011; Suzuki et al. 2005; Tran et al. 2004), but the enzymatic properties of the products of the putative *larA* (AO090023000264) and *lxrA* (AO090038000426) genes have not been reported.

To determine the basic properties of the PCP enzymes in *A. oryzae*, the enzymes were produced in *E. coli* as His-tagged proteins and then purified using Ni-NTA agarose (Fig. S1). Despite exhaustive trials using various strains, LadA was not successfully produced in *E. coli*. The substrate specificities and coenzyme preferences of the PCP enzymes, except for LadA, are shown in Fig. 1 and the kinetic parameters of the enzymes are shown in Table S7. Our results confirm that every gene examined in this study encodes an enzyme that is involved in pentose catabolism. Interestingly, LxrA showed high reactivity to both L-xylulose and D-xylulose, which has not been reported. Although a definitive conclusion cannot be drawn without more precise studies, this observation suggests that LxrA may catalyze conversion of xylitol to D-xylulose as does XdhA.

Effect of *xlnR* and *araR* deletion on the expression of the pentose catabolic genes

Genes that encode enzymes in the pentose catabolic pathway are regulated by XlnR and AraR in *A. niger* and *A. nidulans* (Battaglia et al. 2011a, b; Hasper et al. 2000). In *A. oryzae*, the *xyrA*, *lxrA*, *ladA*, and *xdhA* genes are regulated by XlnR (Noguchi et al. 2009). Here we examined further the roles of XlnR and AraR in regulating the expression of the PCP genes.

To determine how dependent expression of those genes is on XlnR and AraR in *A. oryzae*, Δ*xlnR*, Δ*araR*, and Δ*araR*/Δ*xlnR* mutants were constructed. Growth of those deletion mutants on D-glucose,

D-xylose, L-arabinose, xylitol, and L-arabinitol is shown in Fig. 2. The wild type strain RIB40 (WT) showed naturally poor growth on xylitol, L-arabinose, and L-arabinitol compared to on D-glucose and D-xylose. Decreased growth of the Δ*xlnR* mutant was observed on D-xylose and slightly on xylitol compared to WT, and the Δ*araR* mutant exhibited slightly decreased growth on D-xylose, L-arabinose, and xylitol. The Δ*araR*/Δ*xlnR* mutant hardly grew on D-xylose but weakly grew on L-arabinose, indicating that *A. oryzae* has a system to catabolize L-arabinose in a XlnR/AraR-independent manner. The *araR* and/or *xlnR* deletion did not affect growth on L-arabinitol.

The effect of deleting *xlnR* and *araR* on induction of the PCP genes by commercial D-xylose and L-arabinose preparations was investigated by qRT-PCR analysis (Fig. 3). It should be noted that, although 100% pure D-xylose (based on the certificate of analysis by the manufacturer) was used in this study, commercial preparations of L-arabinose may contain small amounts of the other sugar including D-xylose (Noguchi et al. 2011). Therefore, even if elevated transcription is observed by the L-arabinose preparation, it does not directly mean that L-arabinose is the inducing molecules. Both the D-xylose and L-arabinose preparations induced expression of *xyrA* in the wild type (RIB40) and parent (Pt) strains, with the D-xylose preparation inducing higher expression. Deletion of *xlnR* decreased induction, while deletion of *araR* did not affect induction. Deletion of both *xlnR* and *araR* completely abolished induction. These results indicate that XlnR is the primary mediator of induction of *xyrA* expression, and that AraR can partially substitute for XlnR in a Δ*xlnR* mutant.

Regulation of *larA* was simpler. Its induction was abolished in the Δ*araR* mutant and in the Δ*araR*/Δ*xlnR* mutant, while deletion of *xlnR* did not affect its induction. This finding suggests that XlnR contributes little to regulation of *larA*. Both the D-xylose and L-arabinose preparations induced AraR-dependent expression of *larA*, as *xyrA*, to the same level in the Δ*xlnR* strain. Deletion of *araR* and/or *xlnR* also differentially affected expression of *ladA*, *lxrA*, *xdhA*, and *xkiA* (putative

D-xylulose kinase, AO090020000603). Expression profiles of *ladA* and *xdhA* were similar. The D-xylose and L-arabinose preparations induced comparable levels of expression in the wild-type and Δ*xlnR* mutant strains, while induction was abolished in the Δ*xlnR*Δ*araR* mutant. In the Δ*araR* mutant, little to no induction by the L-arabinose preparation was observed, while weak induction by the D-xylose preparation was observed. In the case of *lxrA*, expression decreased with deletion of either *araR* or *xlnR* and was abolished by simultaneous deletion of both. Induction of *xkiA* by the D-xylose preparation was stronger than that by the L-arabinose preparation, as observed for *xyrA*, but was not significantly affected by deletion of *xlnR*. Induction of *xkiA* decreased with deletion of *araR* and was abolished by deletion of both *xlnR* and *araR*.

Consequently, both XlnR and AraR are involved in inducing expression of these genes, except *larA*, since induction of their expression by the D-xylose and L-arabinose preparations was only abolished in the Δ*araR*/Δ*xlnR* mutant. Additionally, deletion of *araR* caused a more significant decrease in the expression of the genes, except for *xyrA*, than did deletion of *xlnR*. Surprisingly, expression of *lxrA* was extremely low compared to that of the other genes, even in the wild type strain. Considering that expression of *lxrA* in *A. oryzae* IFO4206, which grows normally on L-arabinose, was approximately 100-fold higher than that in RIB40 (unpublished data), a low level of expression of *lxrA* might explain why *A. oryzae* RIB40 grows poorly on L-arabinose (Fig. 2).

Identification of AraR binding sites in the *larA* promoter

Electrophoretic mobility shift assays (EMSA) were employed to identify the binding sites for AraR in the *larA* promoter using the DNA-binding domain of AraR (aa 1-84) fused to His_6 at its C terminus (AraR84-H). AraR84-H was expressed in *E. coli* and purified via affinity column chromatography using Ni-NTA agarose (Fig. S1). AraR84-H bound to the DNA fragment covering -379 to +22 of *larA*, but did not bind to its upstream region (positions -1413 to -333) (data not

shown). The region spanning positions -379 to +22 was divided into four DNA fragments that were used as probes for EMSA. A single shifted band appeared for probe F (positions -295 to -181) while two shifted bands were observed for probe G (positions -192 to -80), suggesting that probe F contained one binding site while probe G contained two (Fig. 4a and b). Examination of the DNA sequences of the F and G fragments revealed that they shared three similar short sequences comprising eight bases each: CGGGTAAT at positions -274 to -281 in the F fragment (site 1), and CGGATAAT at positions -180 to -172 and CGGAAAAT at positions -153 to -161 in the G fragment (sites 2 and 3). The conserved CGG triplet at each site was mutated to determine whether it affects binding of AraR84-H. Mutation of the CGG triplet in site 1 of the F fragment (Fm) completely abolished binding by AraR84-H (Fig. 4c). Mutation of the CGG triplet in either one of the two binding sites (sites 2 and 3) in the G fragment (G1m and G2m) caused the slower migrating shifted band, but not the faster migrating shifted band, to disappear (Fig. 4d). These results indicate that the CGG triplets, or a portion of them, are essential to AraR84-H binding. The significantly weaker intensity of the shifted band for the G1m probe, which had the mutation of CGG in site 2 at positions -180 to -178, compared to that for the G2m probe indicates that site 2 has the stronger affinity to AraR84-H than does site 3.

The core region required for AraR84-H binding was identified by 3-bp mutational scanning. The wild-type probe fragment (Fn) covered positions -296 to -258 of the *larA* promoter. As shown in Fig. 5a, mutations outside of the eight-base sequence CGGGTAAT did not affect binding by AraR84-H, while mutations within the eight-base sequence abolished binding by AraR84-H. To evaluate the contribution of each base in the core region to the binding affinity of AraR84-H, a single-base mutational analysis was performed (Fig. 5b). Each base change at each position affected binding by AraR84-H differently depending on the base exchanged. At position +1, C was the most preferred base followed by A, and G bases at positions +2 and +3 were essential. G was most preferred at

position +4, followed by T, while T was most preferred at position +5, followed by G. At positions +6 and +7, A was preferred most, followed by T and G. At position +8, T and A were comparably preferred, followed by G. It should be noted that C was not acceptable at any positions except +1.

Identification of AraR and XlnR binding sites in the *xyrA* promoters

As shown in Fig. 3, *xyrA* is regulated by both XlnR and AraR. Only a single putative AraR binding site was observed at positions -313 to -320 (CGGGTAAA) on the non-coding strand within the upstream 1000 bases (Fig. 6a). AraR84-H bound to the probe spanning the region from positions -398 to -1, which contains the predicted binding site (probe C), but not to those spanning positions -1000 to -650 and positions -675 to -377 (probes A and B) (Fig. 6b). AraR84-H bound to the shorter probe spanning positions -335 to -298 and mutation of CGG triplet in the CGGGTAAA sequence of the probe abolished formation of the DNA-protein complex, thus confirming that AraR84-H bound to this sequence (Fig. 6c).

Aspergillus XlnR is reported to bind to three similar sequences, GGCTAAA, GGCTAGA, and GGCTGAA (de Vries, et al. 2002, Marui, et al. 2002b, van Peij, et al. 1998b). The promoter region of *xyrA* (positions -1000 to -1) contains one GGCTGA (positions -777 to -782) and two GGCTAA sequences (positions -296 to -301 and positions -264 to -269) on the non-coding strand. XlnR binding sites in the *xyrA* promoter were determined using the DNA-binding domain of XlnR (aa 1-183) fused to a FLAG tag at its N-terminus (F-XlnR183), which was produced in *E. coli* and partially purified (Fig. S1). F-XlnR183 did not bind to the region spanning positions -1000 to -377 (Fig. 6c, probes A and B), indicating that XlnR either does not bind to or has a weak affinity for the GGCTGA sequence in the probe A. When the region spanning positions -398 to -1 was used as a probe, two shifted bands were observed (Fig. 6d, probe C). To examine whether XlnR binds to the putative binding sites at positions -296 to -301 (site X1) and positions -264 to -269 (site X2) within

the probe, further EMSA analyses were performed using wild-type and mutant probes that covered positions -329 to -255, which carries a putative AraR binding site at positions -313 to -320 (Fig. 6e). Mutation of the GGC sequence at site X1 to a TTA sequence caused the slower migrating band to disappear, but the same substitution at site X2 site did not affect binding of F-XlnR183 (Fig. 6f). To our surprise, substituting the GGG sequence to a TTT sequence in the AraR binding site caused the slower migrating band to disappear, indicating that F-XlnR183 bound to the overlapping sequence with the AraR binding site. Consequently, there are two XlnR binding sites in the *xyrA* promoter; one is specific to XlnR and the other can be bound by both XlnR and AraR.

Re-examination of the recognition sequences of XlnR

Because F-XlnR183 appears to bind to a region of the *xyrA* promoter that overlaps with the binding site for AraR (CGGGTAAA), we investigated whether it also binds to the other AraR binding sites (CGGGTAAT, CGGATAAT, and CGGAAAAT) present in the *larA* promoter. Considering that expression of *larA* appears to be XlnR-independent, we did not expect that XlnR would bind at these sites. F-XlnR183 did not bind to the upstream region of *larA* spanning positions -1413 to -333, which contains a single putative binding site for XlnR at positions -785 to -790 (GGCTAG) (data not shown). However, when the fragments E through H (Fig. 4a) were used as probes, F-XlnR183 bound to fragments F and G, which contain the AraR binding consensus (Fig. 7a). Furthermore, F-XlnR183 bound to probe Fn (Fig. 5a), a short probe a AraR binding site, and binding was abolished by three-base mutations within the consensus sequence of the AraR binding site (Fig. 7b). This result is consistent with the result obtained from to the binding analysis using AraR84-H (Fig. 5a), implying that AraR and XlnR can share the sequence CGGGTAAT as a binding site. Regarding the G fragment, F-XlnR183 bound strongly to the G1 probe (positions -190 to -162) and weakly to G2 (positions -173 to -145) (Fig. 7c). The G1 probe contained one AraR binding site

(CGGATAAT) at positions -180 to -173 and a related sequence AGGATAAT at positions -190 to -183. Mutation of the AraR binding sequence abolished binding by F-XlnR183 (Fig. 7d, Amt), while mutation of the related sequence did not affect binding by F-XlnR183 (Fig. 7d, Xm). Consequently, the binding sites for F-XlnR183 in probe G appear to be identical to those for the AraR.

Using the probes shown in Fig. 5b, we examined base preference of XlnR (Fig. 7e). The CGG triplet at positions +1 to +3 and the TA sequence at positions +5 to +6 were essential for binding by XlnR. No base preference was observed at position +4. At position +7, A was nearly essential, though T was also acceptable. Position +8 can be either T or A. Thus, the most preferred binding sequence for XlnR identified in this study is CGGNTAAW.

AraR and XlnR binding sites in the promoters of other PCP genes.

The binding sites for AraR and XlnR in the promoters of the other genes encoding enzymes in the PCP were identified by EMSA (Figs. S2 to S5). The results are summarized in Table 1. In those promoters, the preferred binding site for AraR appears to have the sequence CGGDTAAW. According the mutational analysis shown in Fig. 5, modest binding by AraR was observed even after mutations were introduced at position +1, +5, +6, or +7, which suggests that binding by AraR may also occur at sites slightly different from the above consensus sequence. The results of this search for XlnR binding sites in the promoters of genes encoding enzymes in the PCP confirm our proposal that the consensus sequence of binding sites for XlnR is CGGNTAAW. A notable exception occurs in the binding site for XlnR in the *ladA* promoter, the sequence of which is TGGCTAAA.

All the promoters possessed at least one site that was bound by both AraR and XlnR. One XlnR-specific binding sites was identified in the *xyrA* promoter and two were identified in the *ladA* promoter (Figs. 6 and S2). At all these sites, the base at position +4 was C, which precluded binding by AraR.

Competitive DNA binding by AraR and XlnR

Given that AraR and XlnR share binding sites in the promoters of genes encoding enzymes in the PCP, the question arises as to whether they can simultaneously bind to the same site or compete with each other for binding. To answer this question, a DNA binding analysis was performed using a fixed amount of F-XlnR183 and increasing amounts of Ara84-H. As shown in Fig. 8, F-XlnR183 and Ara84-H competed for the binding to the binding site in the *xdhA* promoter (positions -416 to -409) because binding by F-XlnR183 decreased as the amount of Ara84-H present increased. Simultaneous binding, which slows the migration of a DNA-protein complex, was not observed. Expression of *xdhA* in the presence of L-arabinose is significantly enhanced by deletion of *xlnR* (Fig. 3). This enhancement might result from a lack of competitive binding by XlnR. Consequently, AraR and XlnR appear to compete for DNA binding due to the sharing of the binding site. Such competitive binding was also observed, but not obvious, at sites in the *xyrA* and *larA* promoters (Fig. S6).

Re-examination of XlnR binding sites on the xylanase promoters

A. oryzae XlnR is reported to bind strongly to the GGCTAAA sequence in the promoter of *xynF1* (xylanase F1), the preceding base of which is A (Marui, et al. 2002b). However, F-XlnR183 did not bind to the AGGCTAAA sequence in the A fragment of the *xyrA* promoter (Fig. 6a and d). Additionally, a single-base mutational analysis (Fig. 7d) revealed that the presence of an A at position +1 precludes binding by F-XlnR183. These contradictory results suggest that the bases surrounding the core binding sequence affect the preference of XlnR for the base at the +1 position.

Examination of binding by F-XlnR183 to the *xyrA*, *xdhA*, *xynF1*, and *xynG2* promoters revealed that, although the probes used were of almost identical lengths, protein-DNA complexes formed

from the *xynF1*-X1 and *xynG2*-X1 probes migrated more slowly than those formed from the *xyrA*-X1, *xdhA*-A2, and *xynG2*-X3 probes, while the *xynG2*-X2 probe produced two shifted bands that correspond to the slower and the faster migrating bands (Fig. 9b). These results suggest that the slower migrating bands are protein-DNA complexes formed from dimerized F-XlnR183 while the faster migrating bands are protein-DNA complexes formed from F-XlnR183 monomers. This conclusion is consistent with the presence of palindromic sequences (underlined in Fig. 9a) that might mediate binding of F-XlnR183 dimer. Mutations that disrupt the symmetry of those palindromic sequences led to a defect in binding of the *xynF1*-X1 site by F-XlnR183, and replacing the A at the +1 position with a C resulted in the appearance of the slower migrating band (Fig. 9b). Furthermore, a G to C exchange at the +3 position, which corresponds to the base in the middle of the palindromic sequence, did not affect migration of the DNA-protein complex. Consequently, XlnR preferentially binds to TTAGSCTAA as a dimer and to CGGNTAAW as a monomer.

Discussion

The transcription factors AraR and XlnR in *Aspergillus* are close homologs that must have emerged from a gene duplication event that is unique to *Eurotiomycetes*, since *Sordariomycetes* like *N. crassa* and *T. reesei* (*H. jecorina*) do not possess a counterpart of AraR (Huberman et al. 2016; Tani et al. 2014). AraR and XlnR have overlapping functions as both regulate genes encoding enzymes in the PCPs of *A. niger* and *A. nidulans* (Kowalczyk et al. 2014). Though they look like isoforms that have nearly identical physiological functions, they actually have different functions because they differentially regulate many genes, including glycoside hydrolase and transporter genes, in *A. nidulans* and *A. niger* (Battaglia et al. 2011a, b; Battaglia et al. 2014; de Souza et al. 2013; de Vries et al. 2011; Kowalczyk et al. 2015). These observations suggest that divergent evolution

following the gene duplication event that created AraR and XlnR led to complex and possibly fine-tuned regulatory mechanisms in *Eurotiomycetes*.

In this paper, we report our efforts to examine differences between AraR and XlnR of *A. oryzae*, particularly in their responses to D-xylose and L-arabinose and in their recognition sequences. XlnR appears to prefer the D-xylose preparation to the L-arabinose for induction because the former induced higher expression of targets genes encoding enzymes in the PCP than did the latter in the Δ*araR* mutant; D-xylose induced 1.7- to 3.1-fold more expression of *xyrA*, *ladA*, *lxrA*, and *xdhA* than L-arabinose (Fig. 3). Additionally, D-xylose induced ten-fold more expression of *xkiA* than L-arabinose, possible reasons for which is described below. These results are consistent with the observation that D-xylose leads to phosphorylation of XlnR at a lower concentration (0.1 mM) than does L-arabinose (≥10 mM) (Noguchi et al. 2011). Because the commercial L-arabionse preparation contain minor but non-negligible amount of D-xylose, it is still not clear if L-arabinose itself induces XlnR-dependent gene expression. In addition to inducing XlnR-regulated genes, D-xylose can also repress XlnR-regulated genes via CreA-dependent carbon catabolite repression when it is the only available carbon source (de Vries et al. 1999). Therefore, XlnR's preference for D-xylose probably is more obvious at lower concentrations of D-xylose.

AraR induced expression of *xyrA*, *larA*, *ladA*, *lxrA*, and *xdhA* in response to the L-arabinose preparation to nearly the same degree to which it did in response to the D-xylose preparation in the Δ*xlnR* mutant (Fig. 3), except in the case of *xkiA*, the levels of induced expression of which differed by 2.2-fold. However, considering that a corresponding ten-fold difference was observed in the Δ*araR* mutant, AraR appears to have a weaker preference for the D-xylose preparation than XlnR does. Why expression of *xkiA* is highly induced by the D-xylose preparation even in the Δ*xlnR* mutant is not clear. However, XlnR/AraR-independent induction of expression of *xkiA* by the D-xylose preparation might be involved, especially given that the D-xylose preparation induced

expression of *xkiA* 2.5-fold more than the L-arabinose preparation did, although not statistically significantly, in the Δ*araR*/Δ*xlnR* mutant (Fig. 3).

XlnR appears to regulate all genes encoding enzymes in the PCP except *larA*. However, because XlnR and AraA are partially redundant, only expression of *xyrA* decreased in the Δ*xlnR* mutant. This observation suggests that identification of XlnR-regulated genes requires the absence of *araR*. In a previous report, we proposed, based on DNA microarray analysis, that 75 genes are regulated by XlnR, including all of the genes that encode enzymes in the PCP genes except for *larA* (Noguchi et al. 2009). The results of this study are consistent with that proposal. This success may be attributable to employment of the XlnR-overexpressed and XlnR-disrupted strains for the comparison, since such overexpression of XlnR greatly enhanced expression of XlnR-regulated genes.

The recognition sequences that AraR and XlnR bind are nearly identical. We found that AraR binds to the CGGDTAAW sequences in the promoters of genes encoding enzymes in the PCP, while XlnR binds to CGGNTAAW sequences in those promoters. Thus, XlnR should bind to all AraR binding sites. Indeed, we found binding sites for which AraR and XlnR compete *in vitro* in the *xdhA* *xyrA*, and *larA* promoters (Figs. 9 and S6). Though XlnR-specific binding sites were identified in the *xyrA* and *ladA* promoters, no AraR-specific sites were found. Based on the results of mutational analysis of the CGGGTAAT sequence in the *larA* promoter (Figs. 5 and 7), the consensus motif in the recognition sequence of AraR is looser than that of XlnR. Therefore, it is possible that AraR binds to sequences other than CGGDTAAW to regulate AraR-specific gene expression.

Genome-wide expression profiles of Δ*xlnR*, Δ*araR*, and Δ*xlnR*Δ*araR* mutants of *A. niger* grown on a mixture of L-arabinose and D-xylose and on steam-exploded sugarcane Bagasse suggest that interactions between AraR and XlnR can cause complex patterns of gene regulation (de Souza et al. 2013). Investigating the differences in the sensitivities of AraR and XlnR to inducers and how the similarities in their recognition sequences contribute to their competition to bind DNA may help to

elucidate the complex regulatory mechanisms of hemicellulolytic enzyme genes. Whether the presence of AraR and XlnR helps fungi in the class *Eurotiomycetes* to survive in their respective habitats is unclear.

Current literature reports that XlnR binds to the sequence GGCTAAA and closely related sequences like GGCTGAA and GGCTAGA (van Peij et al. 1998a, b; de Vries et al. 2002; Marui et al. 2002a, b). The results of the precise study that we report here, however, suggest that the recognition sequences of XlnR need to be revised. Those results indicate that XlnR preferentially binds to the sequence CGGNTAAW as a monomer and to the sequence TTAGSCTAA in the promoters of *xynF1* and *xynG2*, and the shorter palindromic sequence TAGSCTA in the *xynG2* promoter, as a dimer (Figs. 8 and 9). The distributions of the binding sites for monomeric XlnR (CGGNTAAW) and dimeric XlnR (TTAGSCTAA or TAGSCTA) in the promoters of 75 putative XlnR target genes (Noguchi et al. 2009) are provided in Table 2. Binding sequences for dimerized XlnR (TTAGSCTAA and TAGSCTA) occur in the promoters of 18 genes, 13 of which encode enzymes that are potentially involved in degradation of hemicellulose and cellulose, such as glycoside hydrolases and accessary enzymes. Binding sites for monomeric XlnR are present in the promoters of 45 genes, 35 of which, including the promoters for genes encoding enzymes in the PCP, have one or more sequences that both AraR and XlnR can bind. These genes might be regulated by both XlnR and AraR. Binding sites for only monomeric XlnR were found in 16 promoters.

Paralogous transcription factors occur in a wide variety of living organisms. Such paralogs must have resulted from gene duplication, after which they can acquire distinct functions. The paralogous transcription factors AraR and XlnR differ only slightly in their recognition sequences. Although some promoters contain XlnR-specific binding sites, many others contain binding sites that can be bound by both XlnR and AraR. The expression of the genes controlled by such promoters is determined by the amount and activity of each transcription factor, as well as interference between

them. Because the affinities of each binding site for AraR and XlnR differ, the effect of such interference differs for each gene. As such, regulation by AraR and XlnR is likely complex. Comprehensive transcriptional analyses and *in vivo* DNA binding studies in the presence of different molar ratios of inducers, along with efforts to uncover underlying molecular mechanisms, will increase our understanding of the complex regulation of genes by AraR and XlnR.

Compliance with Ethical Standards

Funding: This work was partially supported by the Program for Promotion of Basic and Applied Researches for Innovations in Bio-oriented Industry and by the Science and Technology Research Promotion Program for Agriculture, Forestry, Fisheries, and Food Industry.

Conflict of Interest: The authors declare that they have no competing interests.

Ethical approval: This article does not contain any studies with human participants or animals performed by any of the authors.

References

Akada R, Kitagawa T, Kaneko S, Toyonaga D, Ito S, Kakihara Y, Hoshida H, Morimura S, Kondo A, Kida K (2006) PCR-mediated seamless gene deletion and marker recycling in *Saccharomyces cerevisiae*. Yeast 23: 399-405 doi: 10.1002/yea.1365

Battaglia E, Hansen SF, Leendertse A, Madrid S, Mulder H, Nikolaev I, de Vries RP (2011a) Regulation of pentose utilisation by AraR, but not XlnR, differs in *Aspergillus nidulans* and *Aspergillus niger*. Appl Microbiol Biotech 91: 387-397 doi: 10.1007/s00253-011-3242-2

Battaglia E, Klaubauf S, Vallet J, Ribot C, Lebrun MH, de Vries RP (2013) Xlr1 is involved in the transcriptional control of the pentose catabolic pathway, but not hemi-cellulolytic enzymes in *Magnaporthe oryzae*. Fungal Genet Biol 57: 76-84 doi: 10.1016/j.fgb.2013.06.005

Battaglia E, Visser L, Nijssen A, van Veluw GJ, Wosten HA, de Vries RP (2011b) Analysis of regulation of pentose utilisation in *Aspergillus niger* reveals evolutionary adaptations in *Eurotiales*. Stud Mycol 69: 31-38 doi: 10.3114/sim.2011.69.03

Battaglia E, Zhou M, de Vries RP (2014) The transcriptional activators AraR and XlnR from *Aspergillus niger* regulate expression of pentose catabolic and pentose phosphate pathway genes. Res Microbiol 165: 531-540 doi: 10.1016/j.resmic.2014.07.013

Benocci T, Aguilar-Pontes MV, Zhou M, Seiboth B, de Vries RP (2017) Regulators of plant biomass degradation in ascomycetous fungi. Biotechnol Biofuels 10:152 doi:10.1186/s13068-017-0841-x

de Groot MJ, Prathumpai W, Visser J, Ruijter GJ (2005) Metabolic control analysis of *Aspergillus niger* L-arabinose catabolism. Biotechnol Prog 21: 1610-1616 doi: 10.1021/bp050189o

de Souza WR, Maitan-Alfenas GP, de Gouvêa PF, Brown NA, Savoldi M, Battaglia E, Goldman MH, de Vries RP, Goldman GH (2013) The influence of *Aspergillus niger* transcription factors

AraR and XlnR in the gene expression during growth in D-xylose, L-arabinose and steam-exploded sugarcane bagasse. Fungal Genet Biol 60: 29-45 doi: 10.1016/j.fgb.2013.07.007

de Vries RP, Benoit I, Doehlemann G, Kobayashi T, Magnuson JK, Panisko EA, Baker SE, Lebrun MH (2011) Post-genomic approaches to understanding interactions between fungi and their environment. IMA Fungus 2: 81-86 doi: 10.5598/imafungus.2011.02.01.11

de Vries RP, van de Vondervoort PJ, Hendriks L, van de Belt M, Visser J (2002) Regulation of the α-glucuronidase-encoding gene (*aguA*) from *Aspergillus niger*. Mol Gen Genomics 268: 96-102 doi: 10.1007/s00438-002-0729-7

de Vries RP, Visser J, de Graaff LH (1999) CreA modulates the XlnR-induced expression on xylose of *Aspergillus niger* genes involved in xylan degradation. Res Microbiol 150: 281-285 doi: 10.1016/S0923-2508(99)80053-9

Gruben BS, Mäkelä MR, Kowalczyk JE, Zhou M, Benoit-Gelber I, de Vries RP (2017) Expression-based clustering of CAZyme-encoding genes of *Aspergillus niger*. BMC Genomics 18: 900 doi: 10.1186/s12864-017-4164-x

Hasper AA, Visser J, de Graaff LH (2000) The *Aspergillus niger* transcriptional activator XlnR, which is involved in the degradation of the polysaccharides xylan and cellulose, also regulates D-xylose reductase gene expression. Mol Microbiol 36: 193-200 doi: 10.1046/j.1365-2958.2000.01843.x

Huberman LB, Liu J, Qin L, Glass NL (2016) Regulation of the lignocellulolytic response in filamentous fungi. Fung Biol Rev 30: 101-111 doi: 10.1016/j.fbr.2016.06.001

Kaneda J, Sasaki K, Gomi K, Shintani T (2011) Heterologous expression of *Aspergillus oryzae* xylose reductase and xylitol dehydrogenase genes facilitated xylose utilization in the yeast *Saccharomyces cerevisiae*. Biosci Biotechnol Biochem 75: 168-170 doi:

10.1271/bbb.100639

Klaubauf S, Narang HM, Post H, Zhou M, Brunner K, Mach-Aigner AR, Mach RL, Heck AJ, Altelaar AF, de Vries RP (2014) Similar is not the same: differences in the function of the (hemi-)cellulolytic regulator XlnR (Xlr1/Xyr1) in filamentous fungi. Fungal Genet Biol 72: 73-81 doi: 10.1016/j.fgb.2014.07.007

Kowalczyk JE, Benoit I, de Vries RP (2014) Regulation of plant biomass utilization in *Aspergillus*. Adv Appl Microbiol 88: 31-56 doi: 10.1016/B978-0-12-800260-5.00002-4

Kowalczyk JE, Gruben BS, Battaglia E, Wiebenga A, Majoor E, de Vries RP (2015) Genetic interaction of *Aspergillus nidulans galR*, *xlnR* and *araR* in regulating D-galactose and L-arabinose release and catabolism gene expression. PloS One 10: e0143200 doi: 10.1371/journal.pone.0143200

Kunitake E, Hagiwara D, Miyamoto K, Kanamaru K, Kimura M, Kobayashi T (2016) Regulation of genes encoding cellulolytic enzymes by Pal-PacC signaling in *Aspergillus nidulans*. Appl Microbiol Biotechnol 100: 3621-3635 doi: 10.1007/s00253-016-7409-8

Kunitake E, Kobayashi T (2017) Conservation and diversity of the regulators of cellulolytic enzyme genes in Ascomycete fungi. Curr Genet 63: 951-958 10.1007/s00294-017-0695-6

Li N, Kunitake E, Aoyama M, Ogawa M, Kanamaru K, Kimura M, Koyama Y, Kobayashi T (2016) McmA-dependent and -independent regulatory systems governing expression of ClrB-regulated cellulase and hemicellulase genes in *Aspergillus nidulans*. Mol Microbiol 102: 810-826 doi: 10.1111/mmi.13493

Lowry OH, Rosebrough NJ, Farr AL, Randall RJ (1951) Protein measurement with the Folin phenol reagent. J Biol Chem 193: 265-275

Marui J, Kitamoto N, Kato M, Kobayashi T, Tsukagoshi N (2002a) Transcriptional activator, AoXlnR, mediates cellulose-inductive expression of the xylanolytic and cellulolytic genes

in *Aspergillus oryzae*. FEBS Lett 528: 279-282 doi: 10.1016/S0014-5793(02)03328-8

Marui J, Tanaka A, Mimura S, de Graaff LH, Visser J, Kitamoto N, Kato M, Kobayashi T, Tsukagoshi N (2002b) A transcriptional activator, AoXlnR, controls the expression of genes encoding xylanolytic enzymes in *Aspergillus oryzae*. Fungal Genet Biol 35: 157-169 doi: 10.1006/fgbi.2001.1321

Mojzita D, Penttilä M, Richard P (2010a) Identification of an L-arabinose reductase gene in *Aspergillus niger* and its role in L-arabinose catabolism. J Biol Chem 285: 23622-23628 doi: 10.1074/jbc.M110.113399

Mojzita D, Vuoristo K, Koivistoinen OM, Penttilä M, Richard P (2010b) The 'true' L-xylulose reductase of filamentous fungi identified in *Aspergillus niger*. FEBS Lett 584: 3540-3544 doi: 10.1016/j.febslet.2010.06.037

Noguchi Y, Sano M, Kanamaru K, Ko T, Takeuchi M, Kato M, Kobayashi T (2009) Genes regulated by AoXlnR, the xylanolytic and cellulolytic transcriptional regulator, in *Aspergillus oryzae*. Appl Microbiol Biotechnol 85: 141-154 doi: 10.1007/s00253-009-2236-9

Noguchi Y, Tanaka H, Kanamaru K, Kato M, Kobayashi T (2011) Xylose triggers reversible phosphorylation of XlnR, the fungal transcriptional activator of xylanolytic and cellulolytic genes in *Aspergillus oryzae*. Biosci Biotechnol Biochem 75: 953-959 doi: 10.1271/bbb.100923

Ogawa M, Tokuoka M, Jin FJ, Takahashi T, Koyama Y (2010) Genetic analysis of conidiation regulatory pathways in koji-mold *Aspergillus oryzae*. Fungal Genet Biol 47: 10-18 doi: 10.1016/j.fgb.2009.10.004

Suzuki T, Tran LH, Yogo M, Idota O, Kitamoto N, Kawai K, Takamizawa K (2005) Cloning and expression of NAD^+-dependent L-arabinitol 4-dehydrogenase gene (*ladA*) of *Aspergillus oryzae*. J Biosci Bioeng 100: 472-474 doi: 10.1263/jbb.100.472

Takahashi T, Masuda T, Koyama Y (2006a) Enhanced gene targeting frequency in *ku70* and *ku80* disruption mutants of *Aspergillus sojae* and *Aspergillus oryzae*. Mol Genet Genomics 275: 460-470 doi: 10.1007/s00438-006-0104-1

Takahashi T, Masuda T, Koyama Y (2006b) Identification and analysis of Ku70 and Ku80 homologs in the koji molds *Aspergillus sojae* and *Aspergillus oryzae*. Biosci Biotechnol Biochem 70: 135-143 doi: 10.1271/bbb.70.135

Tani S, Kawaguchi T, Kobayashi T (2014) Complex regulation of hydrolytic enzyme genes for cellulosic biomass degradation in filamentous fungi. Appl Microbiol Biotechnol 98: 4829-4837 doi: 10.1007/s00253-014-5707-6

Todd RB, Andrianopoulos A (1997) Evolution of a fungal regulatory gene family: the Zn(II)2Cys6 binuclear cluster DNA binding motif. Fungal Genet Biol 21: 388-405 doi: 10.1006/fgbi.1997.0993

Tran LH, Kitamoto N, Kawai K, Takamizawa K, Suzuki T (2004) Cloning and expression of a NAD^+-dependent xylitol dehydrogenase gene (*xdhA*) of *Aspergillus oryzae*. J Biosci Bioeng 97: 419-422 doi: 10.1016/S1389-1723(04)70229-7

van Peij NN, Gielkens MM, de Vries RP, Visser J, de Graaff LH (1998a) The transcriptional activator XlnR regulates both xylanolytic and endoglucanase gene expression in *Aspergillus niger*. Appl Environ Microbiol 64: 3615-3619

van Peij NN, Visser J, de Graaff LH (1998b) Isolation and analysis of *xlnR*, encoding a transcriptional activator co-ordinating xylanolytic expression in *Aspergillus niger*. Mol Microbiol 27: 131-142 doi: 10.1046/j.1365-2958.1998.00666.x

Table 1. AraR and XlnR binding sites on the promoters of the PCP genes.

Gene	Position	Sequence	XlnR/AraR Binding
larA	-274 to -281	CGGGTAAT	XlnR/AraR
	-180 to -173	CGGATAAT	XlnR/AraR
	-153 to -160	CGGAAAAT	XlnR (weak)/AraR (weak)
xyrA	-313 to -320	CGGGTAAA	XlnR/AraR
	-295 to -302	CGGCTAAA	XlnR
ladA	-856 to -849	TGGCTAAA	XlnR
	-834 to -841	CGGTTAAA	XlnR/AraR
	-397 to -404	CGGAGAAT	AraR (weak)
	-320 to -313	CGGCTAAA	XlnR
lxrA	-194 to -187	CGGTTAAA	XlnR (weak)/AraR
xdhA	-438 to -431	CGGTTAAA	XlnR/AraR
	-416 to -409	CGGTTAAA	XlnR/AraR
xkiA	-649 to -642	CGGGTAAA	XlnR/AraR

Table 2. Distribution of the XlnR recognition sequences (CGGNTAAW, TTAGSCTAA, TAGSCTA) on the promoters of XlnR-regulated genes (Noguchi, et al. 2009).

Gene	HXlnR/ ΔXlnR*	Sequence	Position	Protein
AO090701000885	95.5	CGGCTAAA	-139 to -146	Arabinoxylan arabinofuranohydrolase
AO090023000787	77.1	CGGATAAT	-279 to -272	Lytic polysaccharide monooxygenase
AO090103000423	75.3	ATTAGGCTAAA	-167 to -157	Endoxylanase (XynF1)
AO090120000025	69.7	None		MFS transporter
AO090003000553	63.0	None		Hypothetical protein
AO090120000026	61.3	CGGGTAAT	-489 to -496	Endoxylanse (XynG2)
		CTTAGGCTAAA	-652 to -642	
		CATAGGCTAAA	-443 to -453	
AO090001000207	56.1	CGGGTAAA	-160 to -167	Feruloyl esterase (*Ang*
		TTTAGGCTAAG	-837 to -827	FaeA)
		TTTAGGCTATT	-601 to -591	
AO090011000745	55.1	CGGCTAAT	-369 to -376	Acetylxylan esterase
		TCTAGGCTATT	-535 to -525	(AxeA)
AO090005000698	54.2	CGGCTAAA	-132 to -125	Putative β-Xylosidase
		CGGATAAA	-314 to -321	
			-185 to -192	
		TTTAGGCTATA	-143 to -133	
AO090001000208	52.3	CGGTTAAA	-170 to -163	Endoxylanase (XynF3)
AO090003001277	50.2	CGGCTAAA	-190 to -197	MFS transporter
AO090001000069	49.8	CGGATAAA	-738 to -731	MFS transporter (*And*
		ATTAGGCTAGA	-500 to -510	XtrD)
AO090011000140	48.6	CGGCTAAA	-639 to -646	β-Xylosidase (*And*
		CGGGTAAT	-435 to -442	XlnD)
		TTTAGGCTAAA	-605 to -595	
		ATTAGGCTAAA	-527 to -537	
		TCTAGGCTAAA	-165 to -175	
AO090003001341	45.7	CGGATAAT	-436 to -443	Glycoside hydrolase family 5

AO090701000274	42.1	CGGCTAAA	-598 to -605	Glycoside hydrolase
			-457 to -464	family 3
		CCTAGGCTAGA	-730 to -740	
AO090001000111	34.8	CGGATAAA	-267 to -260	Endoxylanase (XynG1)
AO090003000859	33.1	CGGGTAAA	-313 to -320	Xylose reductase (XyrA)
		CGGCTAAA	-295 to -302	
AO090038000439	30.8	CGGCTAAA	-178 to -185	Putative
		AATAGGCTATG	-438 to -448	cellobiohydrolase
AO090011000141	27.7	CGGGTAAT	-440 to -433	Putative exoarabinanase
		CGGCTAAA	-236 to -229	
		TCTAGGCTAAA	-710 to -700	
		ATTAGGCTAAT	-348 to -338	
		TTTAGGCTAAA	-270 to -260	
AO090005000423	24.6	CGGATAAA	-267 to -274	Glycoside hydrolase
				family 5
AO090701000639	23.9	CGGATAAA	-66 to -59	Glycoside hydrolase
				family 31
AO090001000267	20.2	AATAGGCTATT	-202 to -192	Hypothetical protein
AO090005000986	18.7	CGGGTAAT	-547 to -554	β-Xylosidase XylA
AO090020000042	17.9	CGGTTAAA	-226 to -219	Putative
		TTTAGGCTAAT	-146 to -156	aldose-1-epimerase
AO090010000515	17.6	None		Putative
				3-hydroxyisobutylate
				dehydrogenase
AO090701000886	16.4	None		Putative
				arabinofuranosidase
AO090005001169	16.1	None		Fungal transcription
				factor
AO090010000314	15.8	CGGCTAAA	-442 to -435	Endoglucanase (CelB)
		CGGCTAAA	-240 to -233	
AO090012000941	15.5	None		Cellobiohydrolase
				(CelD)
AO090701000884	14.9	None		Putative feruloyl esterase
AO090003000782	14.6	CGGTTAAA	-862 to -869	MFS transporter
		CGGGTAAA	-458 to -465	
			-255 to -262	

AO090701000828	14.4	None		Hypothetical protein
AO090005001078	14.1	CGGTTAAA	-834 to -841	L-arabinitol-4-
		CGGCTAAA	-320 to -313	dehydrogenase (LadA)
		TATAGGCTATT	-628 to -618	
AO090001000649	13.0	CGGCTAAA	-452 to -445	Putative α-xylosidase
AO090038000631	12.7	CGGTTAAA	-438 to -431	Xylitol dehydrogenase
			-416 to -409	(XdhA)
AO090005000768	11.6	CGGATAAT	-330 to -323	Putative α-xylosidase
		TTTAGGCTATA	-377 to -367	
AO090012000591	11.4	CGGGTAAA	-778 to -785	Hypothetical protein
AO090103000087	11.3	CGGGTAAT	-736 to -729	Putative polysaccharide
		CGGTTAAA	-516 to -523	monooxygenase
		ATTAGGCTAAA	-672 to -682	
AO090001000348	11.3	CGGGTAAA	-796 to -759	Cellobiohydrolase
		CGGTTAAA	-795 to -802	(CelC)
AO090001000383	11.0	None		Hypothetical protein
AO090005000337	10.7	ACTAGGCTATG	-519 to -509	Glycoside hydrolase
				family 3
AO090023000401	9.8	CGGGTAAT	-722 to -715	Polygalacturonase
				(PgaB)
AO090012000445	8.8	CGGCTAAA	-851 to -844	β-Galactosidase (*A. niger*
		CGGATAAT	-765 to -758	LacA)
AO090103000268	8.3	None		Glycoside hydrolase
				family 43
AO090005000189	8.2	None		Hypothetical protein
AO090026000127	8.2	CGGCTAAA	-261 to -254	α-Glucuronidase
AO090102000630	8.0	None		Homocysteine synthase
AO090102000040	7.9	CGGATAAA	-174 to -167	Aldo/keto reductase
		CGGGTAAA	-143 to -136	family
AO090103000138	7.5	None		Putative sterol C5
				desaturase
AO090003000905	7.5	CGGATAAT	-393 to -386	Xyloglucan-specific
		ATTAGGCTATA	-556 to -546	endoglucanase
AO090038000426	7.5	CGGTTAAA	-194 to -187	L-xylulose reductase
				(LxrA)
AO090103000426	7.5	CGGATAAA	-85 to -92	Hypothetical protein

AO090038000507	7.4	None		Hypothetical protein
AO090005000518	7.3	CGGATAAA	-360 to -367	Esterase_lipase superfamily
AO090011000944	7.0	None		Fungal transcription factor
AO090011000241	7.0	None		MFS transporter
AO090011000483	6.7	None		MFS transporter
AO090701000345	6.6	CGGCTAAA	-188 to -195	SGNH_hydrolase
			-179 to -186	superfamily (XynE_like)
		TATAGGCTACT	-81 to -71	
AO090026000102	6.5	None		Endoglucanase CelA
AO090005001147	6.5	None		Zinc-containing dehydrogenase family
AO090001000164	6.5	None		Hypothetical protein
AO090010000729	6.3	None		Hypothetical protein
AO090003001484	6.1	CGGCTAAT	-313 to -306	Cell wall protein (*A. nidulans* BinB/PhiA)
		AGTAGGCTACC	-285 to -295	
AO090011000376	6.1	None		Putative metalloreductase
AO090012000034	5.9	CGGTTAAA	-443 to -450	Putative cyclopentanone 1,2-monooxygenase
AO090020000603	5.9	CGGGTAAA	-649 to -642	Putative D-xylulose kinase (XkiA)
AO090026000784	5.8	None		Putative exopolygalacturonase
AO090701000826	5.7	None		Putative polyketide synthase
AO090005000531	5.7	CGGATAAA	-343 to -336	Putative polysaccharide monooxygenase
AO090005001427	5.7	None		Putative glucose-6-phosphate 1-dehydrogenase
AO090003000497	5.5	CGGATAAA	-204 to -211	Putative β-glucosidase
AO090003000018	5.4	None		Putative Hsp30
AO090010000063	5.4	CGGTTAAA		MFS transporter
AO090206000065	5.1	None		Hypothetical protein

AO090003000463	5.1	CGGATAAA	-320 to -327	Hypothetical protein

*Fold increase in the expression of the genes in the XlnR-overproducing strain (HXlnR) as compared to the XlnR disruptant (ΔXlnR) (Noguchi, et al. 2009)).

Figure captions

Fig. 1 Substrate specificities and coenzyme preferences of D-xylose reductase (XyrA), L-arabinose reductase (LarA), L-xylulose reductase (LxrA), and xylitol dehydrogenase (XdhA) in *A. oryzae*. The values are means from three independent experiments. Error bars indicate the standard deviations.

Fig. 2 Growth of strain RIB40 (WT) and the Δ*xlnR* (ΔX), Δ*araR* (ΔR), and Δ*xlnR*Δ*araR* (ΔXA) mutants on various carbon sources. Conidiospores (1×10^4) of strains RIB40 (WT), Δ*xlnR* (ΔX), Δ*araR* (ΔA), and Δ*araR*/Δ*xlnR* (ΔXA) mutants were spotted onto Czapek Dox plates containing 2% of carbon sources and incubated at 30°C for 7 days. a) Typical appearance of the colonies of WT and the mutants on the various carbon sources. b) Colony diameters of WT and the mutants on various carbon sources. The values are means of three independent experiments. Error bars indicate the standard deviations. Different letters above the data bars indicate a significant difference between the tested strains (one-way ANOVA followed by post hoc Tukey's test, $p<0.05$).

Fig. 3 Induction of expression of the pentose catabolic genes by D-xylose and L-arabinose in strains RIB40 (WT) and RkuN16ptr1 (Pt) and in the Δ*xlnR*, Δ*araR*, and Δ*xlnR*/Δ*araR* mutants. Expression levels are shown relative to that of the β-actin gene (*actA*). The values are means from three independent experiments. Error bars indicate the standard deviations. Letters above the bars indicate a significant difference (one-way ANOVA followed by post hoc Tukey's test, $p<0.05$). F = 6 mM D-fructose, X = 6 mM fructose + 25 mM D-xylose, A = 6 mM fructose + 25 mM L-arabinose.

Fig. 4. Identification of AraR binding sites in the *larA* promoter. a) The promoter regions used as probes for EMSA. The numbers represent the base positions, with the translational start site as +1.

The positions of the putative AraR binding sites are shown by the black boxes. b) Results of EMSA to examine binding of AraR84-H to probes E - H. c) The effect of a mutation in the putative AraR binding site in probe F on binding by AraR84-H. d) The effect of mutations in the putative AraR binding sites in probe G on binding by AraR84-H.

Fig. 5 Mutational studies of the AraR binding site in the *larA* F fragment. Effect of mutations were analyzed by EMSA using AraR84-H (0.6 μg). a) Three-base scanning mutagenesis of the AraR binding site and its surrounding sequences. The probe sequences used for the analysis are shown on the left. Base changes are indicated with small letters. Bold letters indicate the core binding sequence. b) The base preferences of AraR at different positions within its recognition sequence. Base changes were introduced at each base in the eight-base core sequence

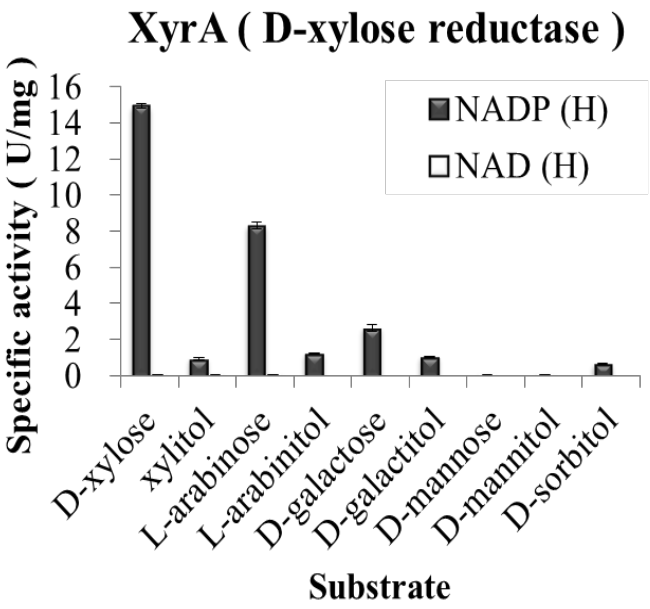
Fig. 6 Identification of AraR and XlnR binding sites in the *xyrA* promoter. a) The promoter regions used as probes in EMSA. The numbers represent the base positions relative to the translational start site as +1. The positions of putative AraR and XlnR binding sites are shown by the black and gray boxes, respectively. b) Results of EMSA to examine binding of AraR84-H to probes A - C. c) The effects of mutations introduced into the putative AraR binding site on binding by AraR84-H. The promoter region spanning positions -335 to -298 was used as the probe. WT = wild type probe. MT = mutant probe. d) Results of EMSA to examine binding of F-XlnR183 to probes A - C. e) The effect of the mutations introduced into putative AraR and XlnR binding sites on binding by F-XlnR183. The probes used for the study are shown on the left. A, X1, and X2 indicate the putative AraR and XlnR binding sites.

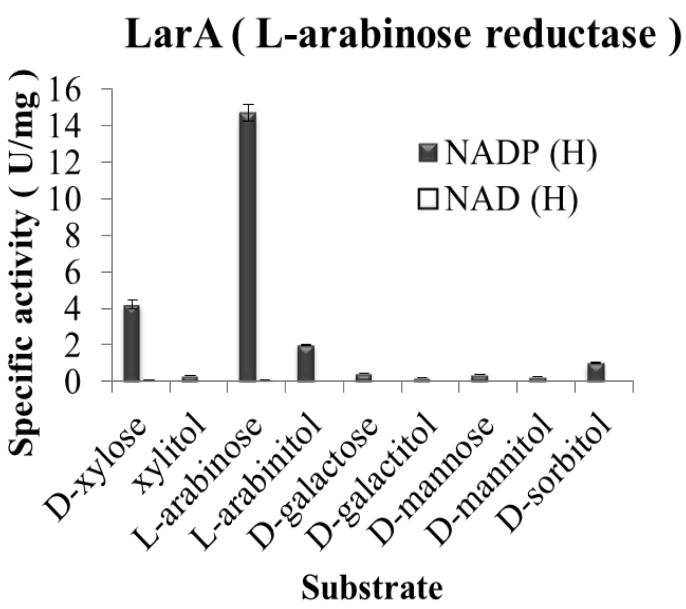
Fig. 7 Identification of XlnR binding sites on the *larA* promoter. a) Results of EMSA to examine

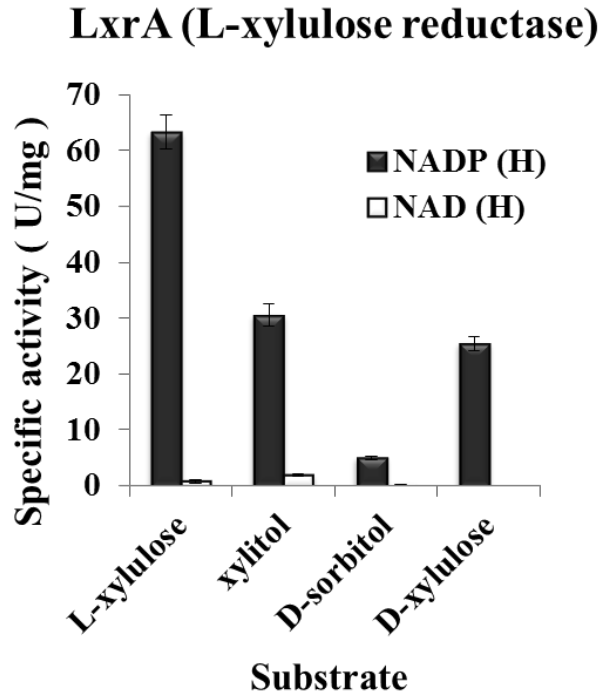
binding of F-XlnR183 to probes E - H from *larA* (Fig. 4a). b) Results of a 3-base scanning mutagenesis performed to determine the core motif of the XlnR recognition sequence. The probes in Fig. 5a were used with F-XlnR183 (0.6 μg). c) Results of EMSA used to identify XlnR binding sites in probe G. The probes used are shown at the top. Small letters in the sequences indicate mutations that were introduced in the probes. Bold letters indicate the possible core recognition motifs. d) Results of EMSA used to determine the base preferences of XlnR at different positions within its DNA recognition sequence. The probes in Fig. 5b were used with F-XlnR183 (0.6 μg)

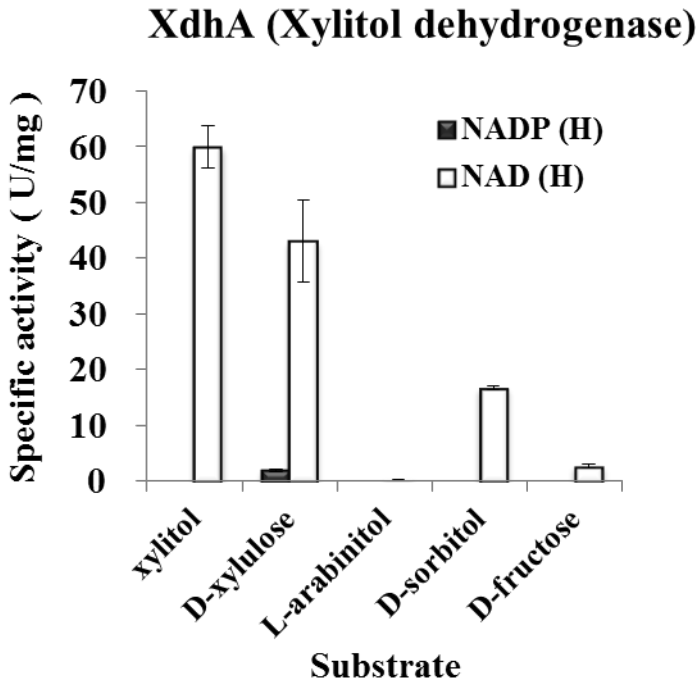
Fig. 8 Competitive DNA binding of AraR and XlnR at the shared binding site in the *xdhA* promoter. The probe sequence used in the study is shown on the top

Fig. 9 Re-examination of XlnR binding sites in the xylanase promoters. a) The probes used in the study are shown at the top. Bold letters indicate identified potential core recognition motifs of XlnR. Palindromic sequences are underlined. b) Mutational analysis of the binding site of dimerized XlnR in the *xynF1* promoter. The binding site and its surrounding sequences are shown at the top. The numbers +1 to +8 corresponds to those of the monomer binding sites shown in Figs. 5 and 7. Palindromic sequence is underlined

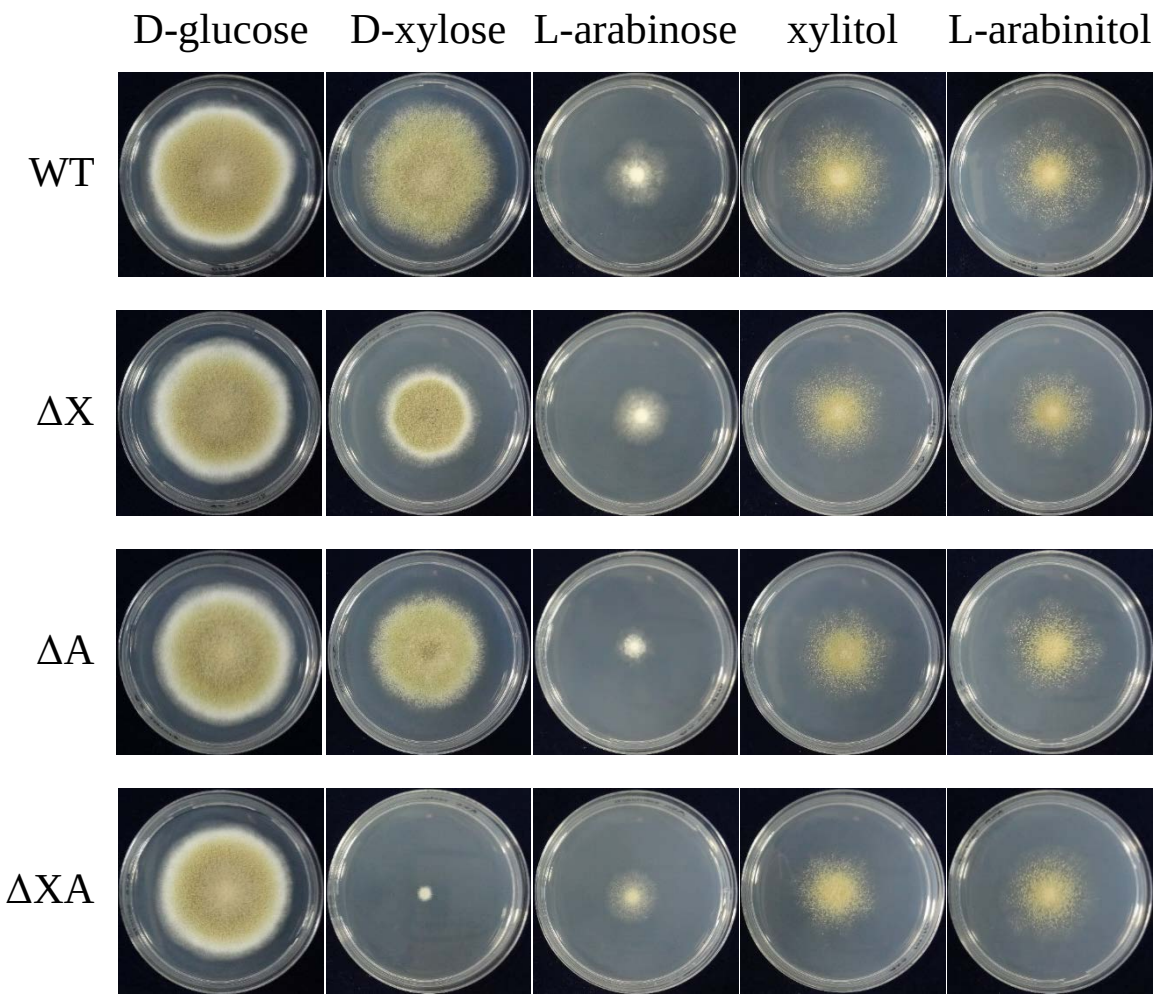

XyrA (D-xylose reductase)
Specific activity (U/mg)
16
14
12
10
8
6
4
2
0
NADP (H)
NAD (H)
D-xylose
xylitol
L-arabinose
L-arabinitol
D-galactose
D-galactitol
D-mannose
D-mannitol
D-sorbitol
Substrate


LarA (L-arabinose reductase)
Specific activity (U/mg)
16
14
12
10
8
6
4
2
0
NADP (H)
NAD (H)
D-xylose
xylitol
L-arabinose
L-arabinitol
D-galactose
D-galactitol
D-mannose
D-mannitol
D-sorbitol
Substrate

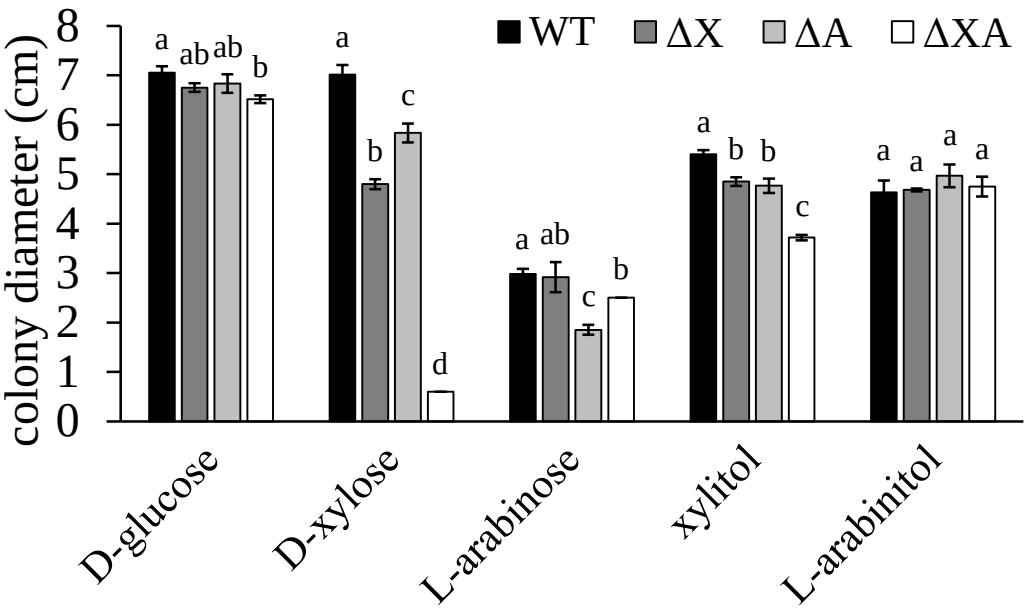

LxrA (L-xylulose reductase)
Specific activity (U/mg)
70
60
50
40
30
20
10
0
NADP (H)
NAD (H)
L-xylulose
xylitol
D-sorbitol
D-xylulose
Substrate

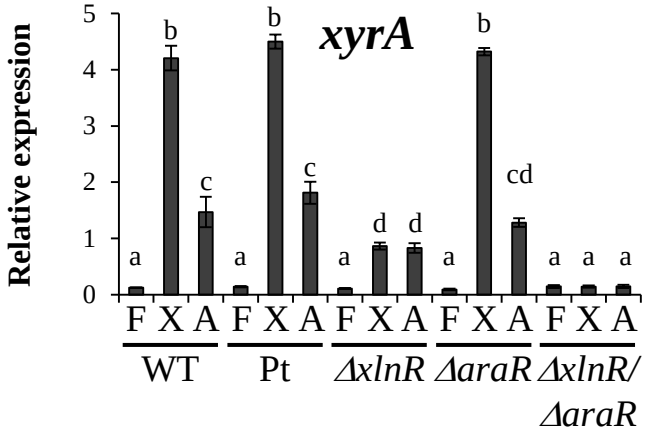

XdhA (Xylitol dehydrogenase)
Specific activity (U/mg)
70
60
50
40
30
20
10
0
NADP (H)
NAD (H)
xylitol
D-xylulose
L-arabinitol
D-sorbitol
D-fructose
Substrate

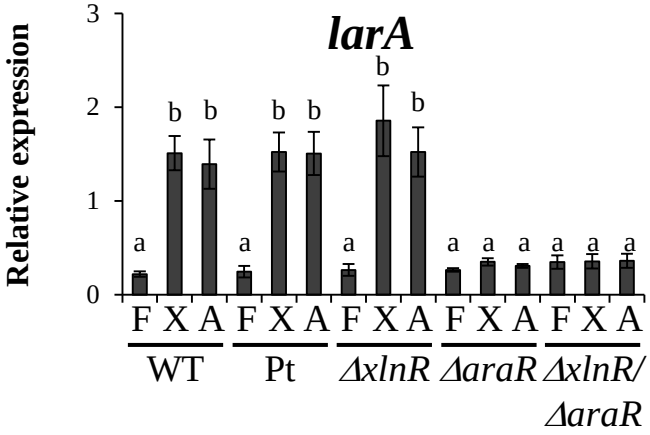
a

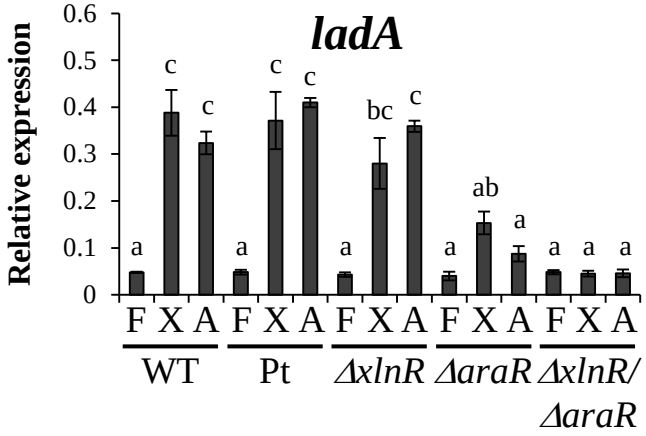

D-glucose
D-xylose
L-arabinose
xylitol
L-arabinitol
WT
ΔX
ΔA
ΔXA

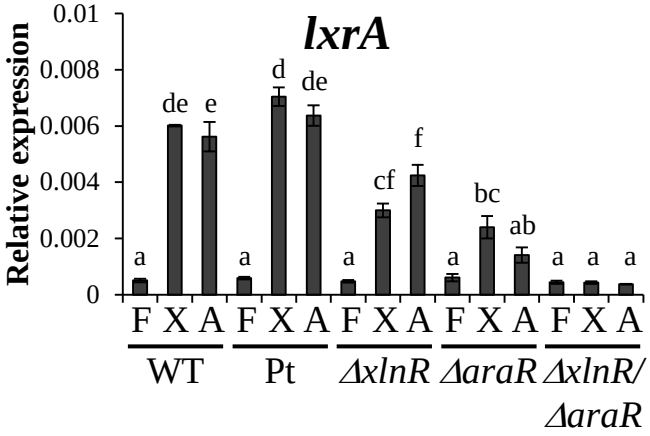
b

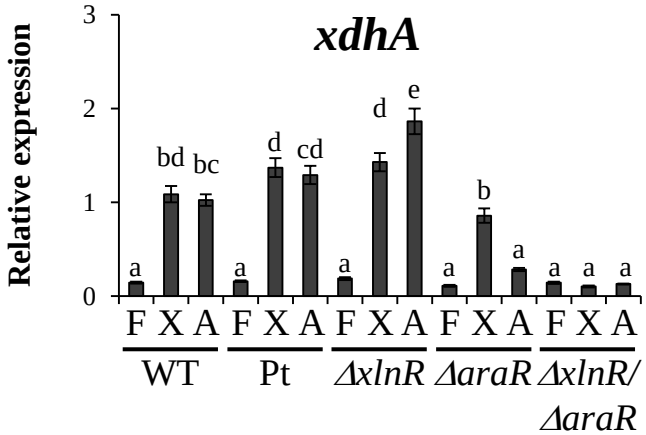

WT
ΔX
ΔA
ΔXA
colony diameter (cm)
0
1
2
3
4
5
6
7
8
a
ab
ab
b
a
b
c
d
a
ab
c
b
a
b
b
c
a
a
a
a
D-glucose
D-xylose
L-arabinose
xylitol
L-arabinitol

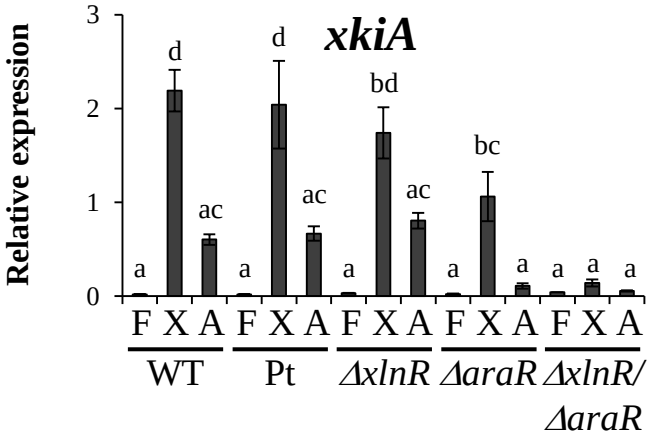

xyrA
Relative expression
F X A F X A F X A F X A F X A
WT Pt ΔxlnR ΔaraR ΔxlnR/ ΔaraR

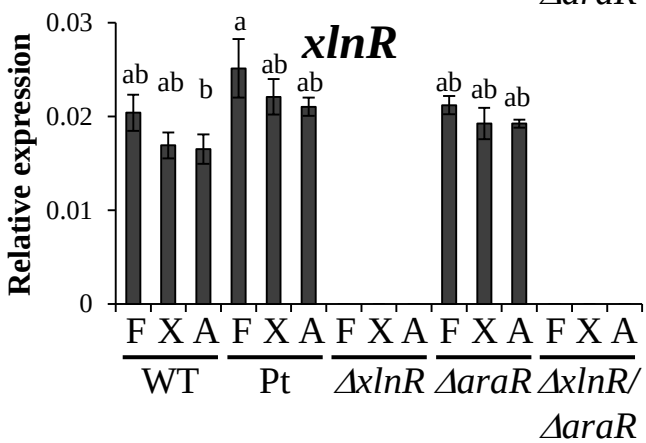

larA
Relative expression
F X A F X A F X A F X A F X A
WT Pt ΔxlnR ΔaraR ΔxlnR/ ΔaraR

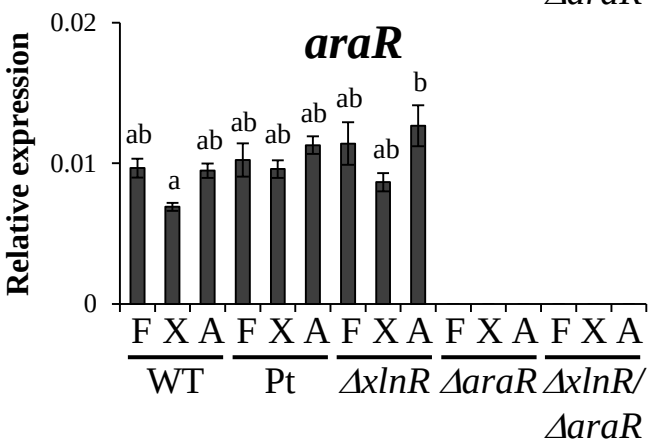

ladA
Relative expression
F X A F X A F X A F X A F X A
WT Pt ΔxlnR ΔaraR ΔxlnR/ ΔaraR


lxrA
Relative expression
F X A F X A F X A F X A F X A
WT Pt ΔxlnR ΔaraR ΔxlnR/ ΔaraR

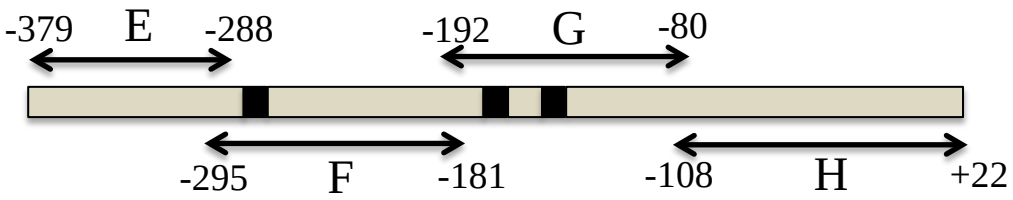

xdhA
Relative expression
F X A F X A F X A F X A F X A
WT Pt ΔxlnR ΔaraR ΔxlnR/ ΔaraR


xkiA
Relative expression
F X A F X A F X A F X A F X A
WT Pt ΔxlnR ΔaraR ΔxlnR/ ΔaraR

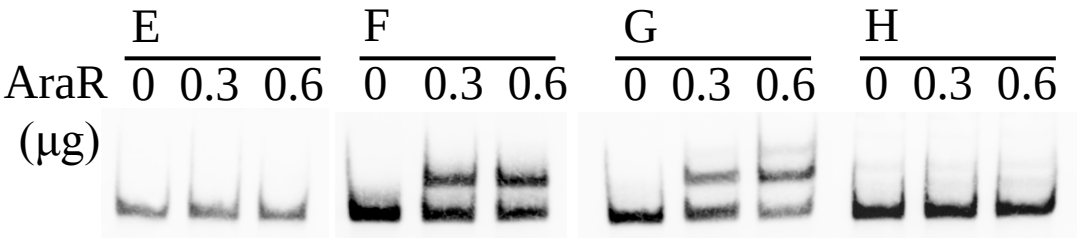

xlnR
Relative expression
F X A F X A F X A F X A F X A
WT Pt ΔxlnR ΔaraR ΔxlnR/ ΔaraR


araR
Relative expression
F X A F X A F X A F X A F X A
WT Pt ΔxlnR ΔaraR ΔxlnR/ ΔaraR

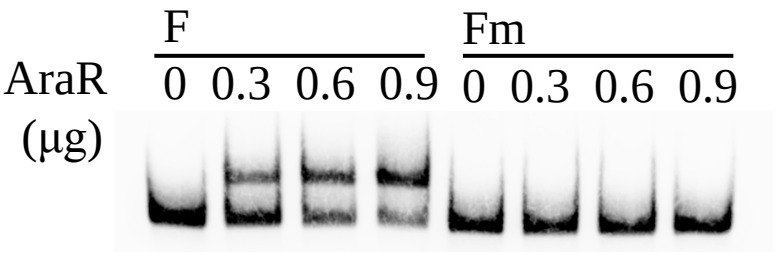
a


-379
E
-288
-192
G
-80
-295
F
-181
-108
H
+22

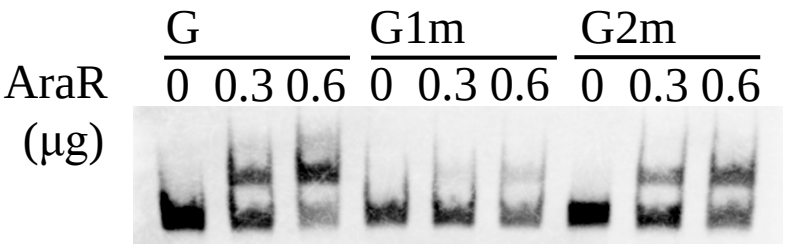
b


E
F
G
H
AraR
(μg)
0 0.3 0.6
0 0.3 0.6
0 0.3 0.6
0 0.3 0.6

c

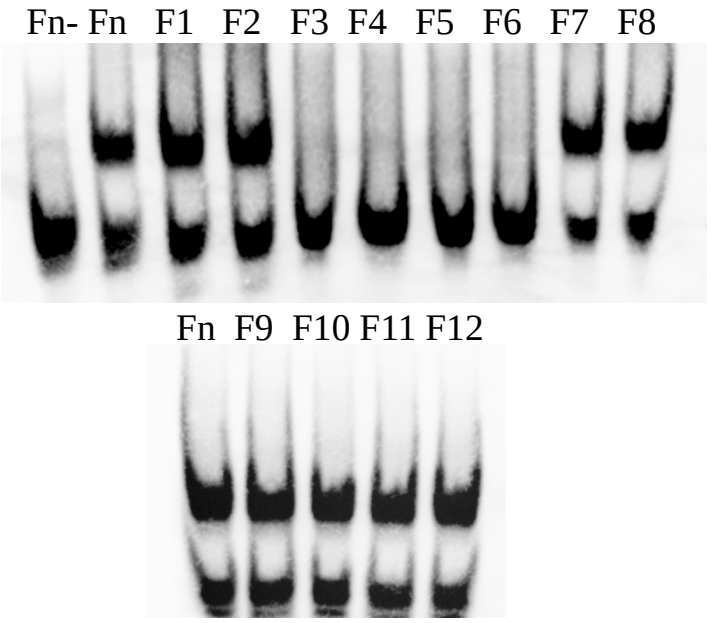

F
Fm
AraR
(μg)
0 0.3 0.6 0.9
0 0.3 0.6 0.9

d

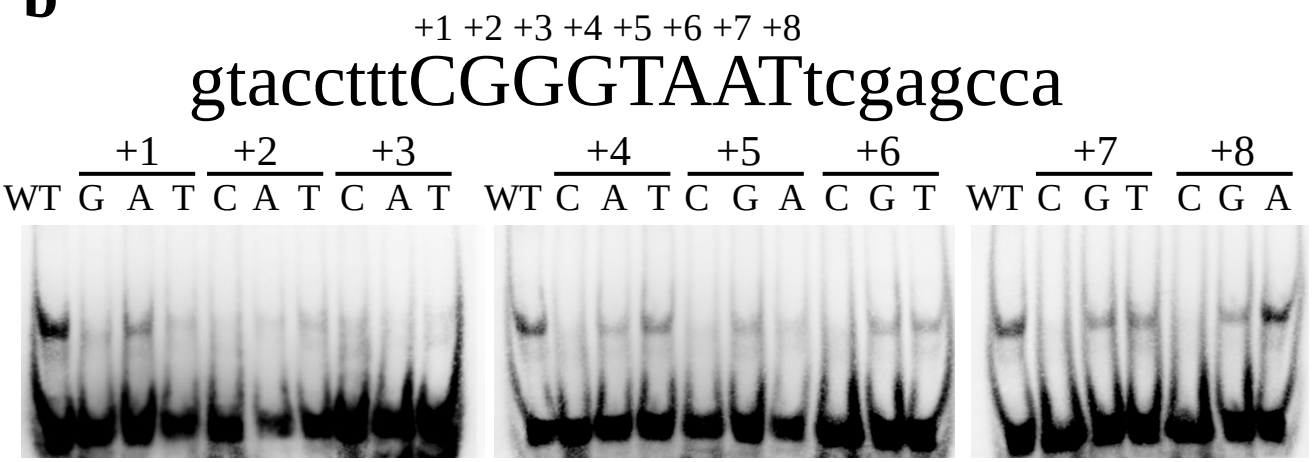

G
G1m
G2m
AraR
(μg)
0 0.3 0.6
0 0.3 0.6
0 0.3 0.6

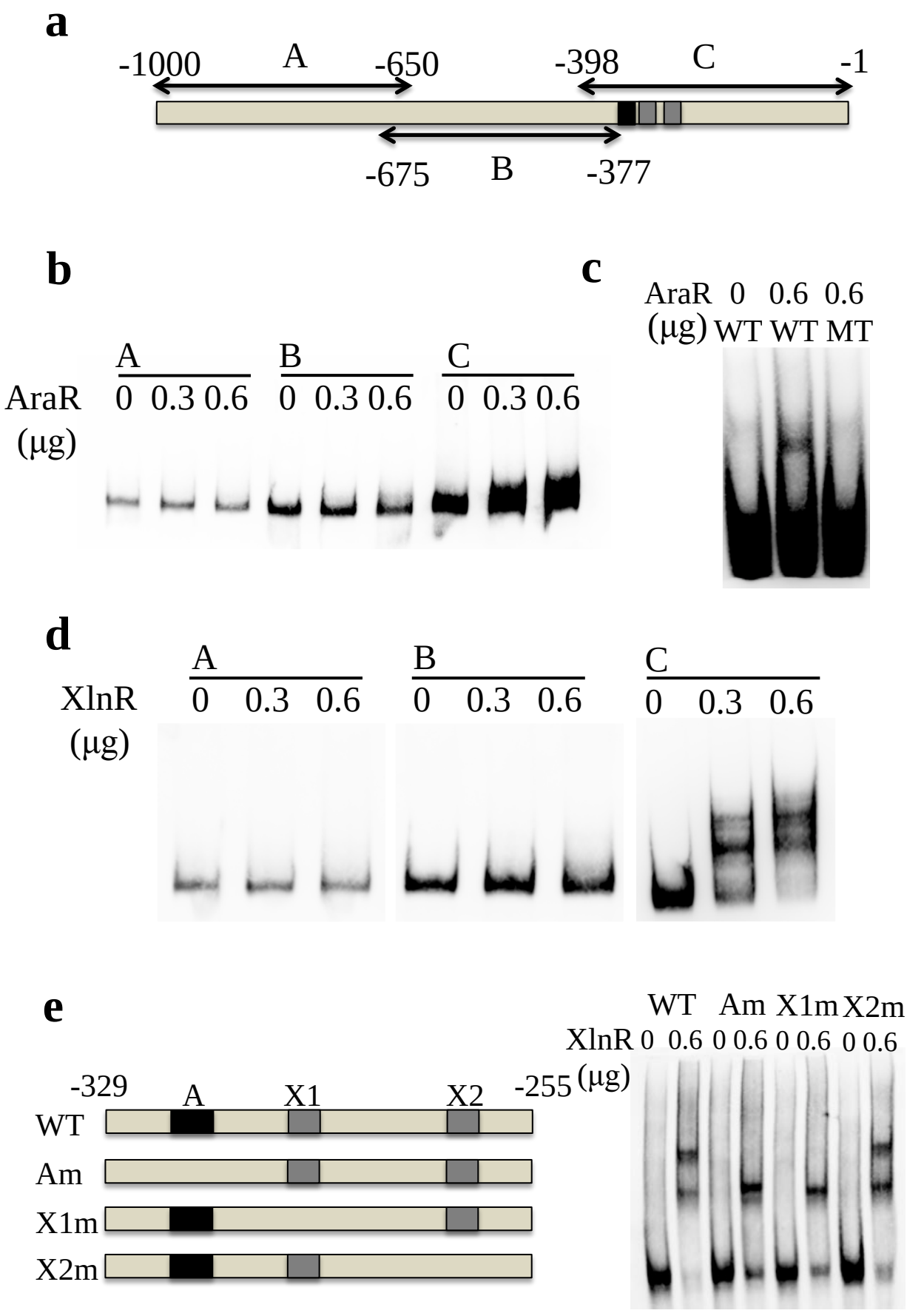
a

```
Fn   GTACCTTTCGGGTAATTCGAGCCA
F1   GTACCTTTCGGGTAATTCGAGaac
F2   GTACCTTTCGGGTAATTCtctCCA
F3   GTACCTTTCGGGTAAggaGAGCCA
F4   GTACCTTTCGGGgccTTCGAGCCA
F5   GTACCTTTCtttTAATTCGAGCCA
F6   GTACCTggaGGGTAATTCGAGCCA
F7   GTAaagTTCGGGTAATTCGAGCCA
F8   tgcCCTTTCGGGTAATTCGAGCCA
F9   GTACCTTTCGGGTAATTatcGCCA
F10  GTACCTTTCGGGTAATgatAGCCA
F11  GTACCgggCGGGTAATTCGAGCCA
F12  GTACaggTCGGGTAATTCGAGCCA
```

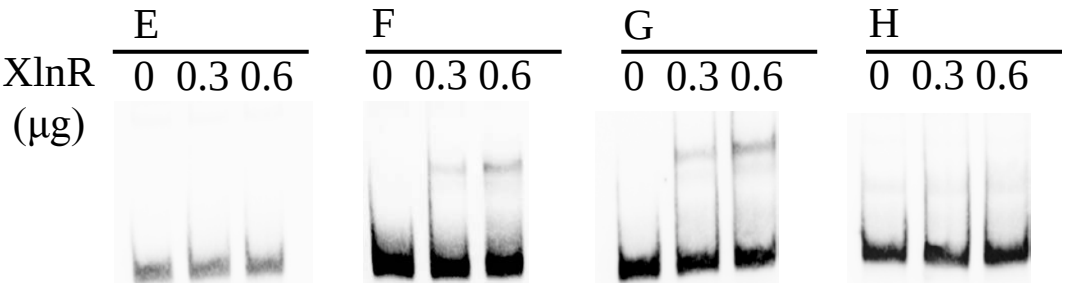


b

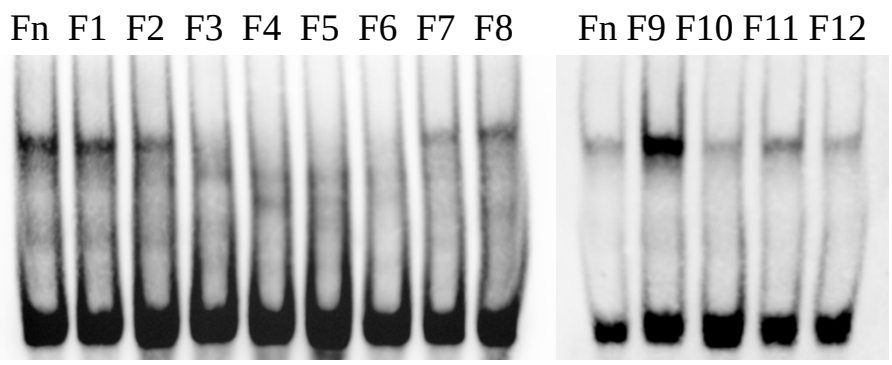



a
-1000
A
-650
-398
C
-1
-675
B
-377
b
A
B
C
AraR
(μg)
0 0.3 0.6 0 0.3 0.6 0 0.3 0.6
c
AraR 0 0.6 0.6
(μg) WT WT MT
d
A
B
C
XlnR
(μg)
0 0.3 0.6 0 0.3 0.6 0 0.3 0.6
e
-329
A
X1
X2
-255
WT
Am
X1m
X2m
WT Am X1m X2m
XlnR 0 0.6 0 0.6 0 0.6 0 0.6
(μg)

a

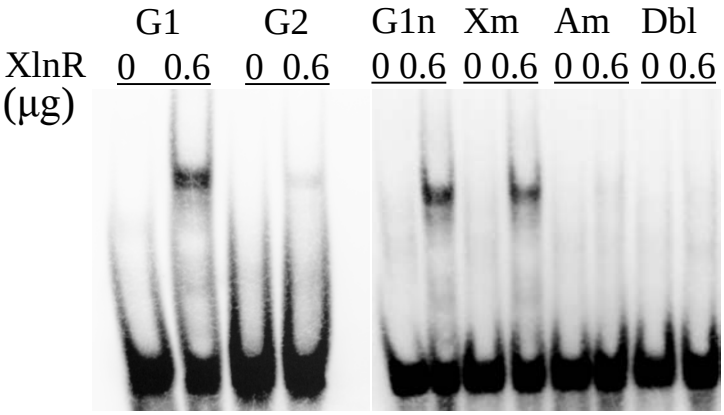


b

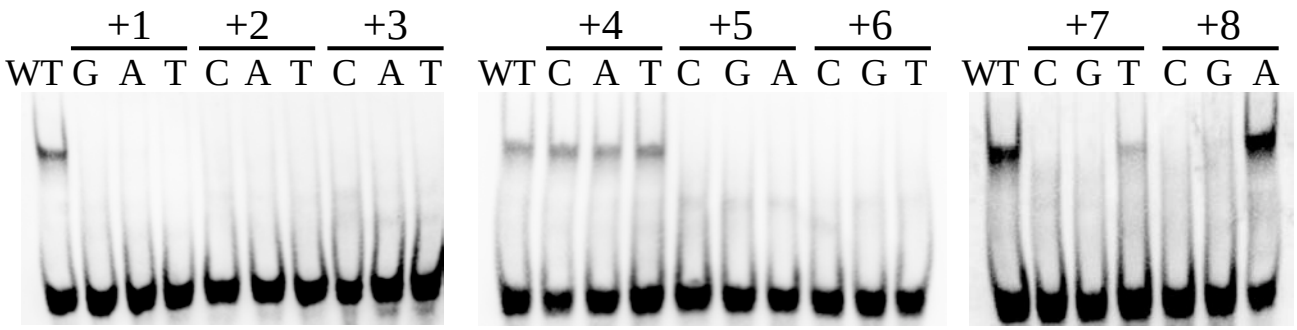


c

G1 CAGGATAATAG**CGGATAAT**TTCAATCCAGCG
G2 CAAGGTCCT**CGGAAAAT**TGCTGGATTGAAAG
G1n CGCGTG**AGGATAAT**AG**CGGATAAT**TTCAATCCAGCG
Xm CGCGTG**A**aat**TAAT**AG**CGGATAAT**TTCAATCCAGCG
Amt CGCGTG**AGGATAAT**AG**C**aat**TAAT**TTCAATCCAGCG
Dbl CGCGTG**A**aat**TAAT**AG**C**aat**TAAT**TTCAATCCAGCG

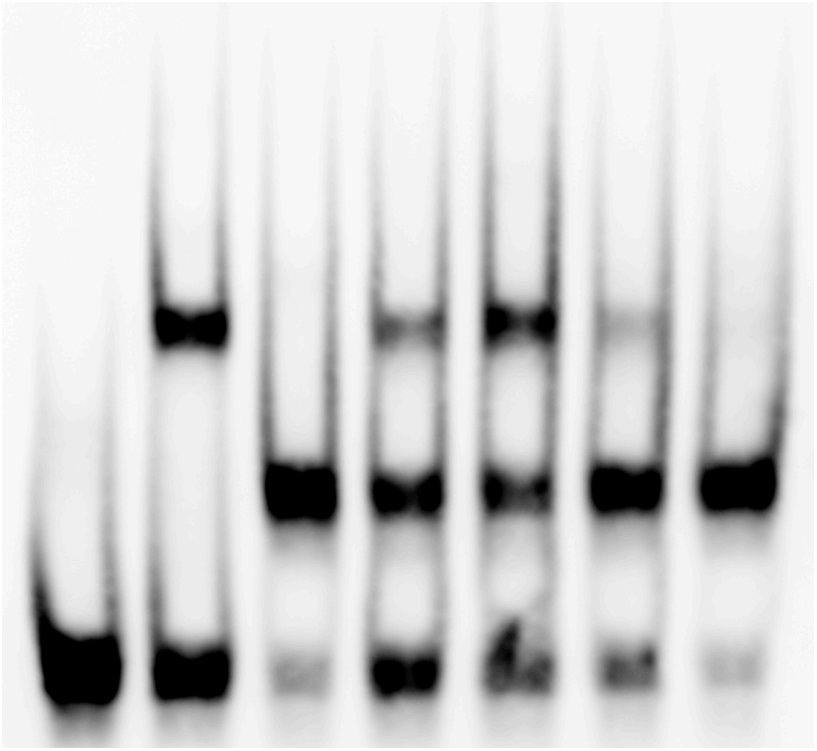


d



xdhA-A2 site (-416 to -409)

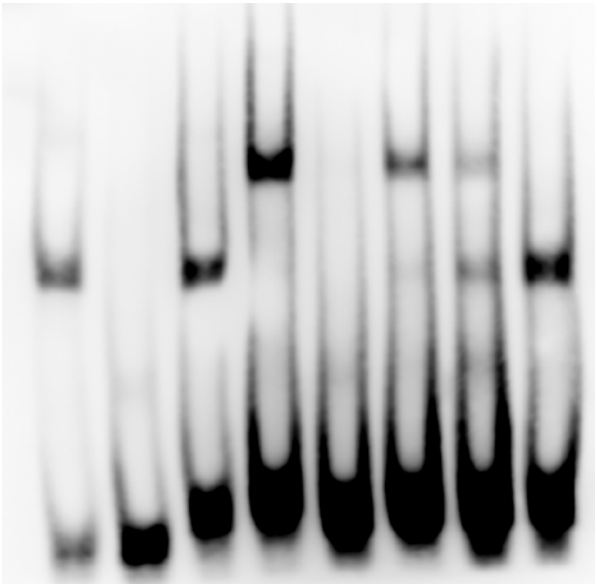
AATGAAAAGAACCAG**CGGTTAAA**TTCCATCTGCTTTAG



F-XlnR183	0	0.6	0	0.6	0.6	0.6	0.6	
AraR84-H	0	0	0.6	0.075	0.15	0.3	0.6	(μg)

a

xyrA-X1	ACAAGATCCCATTC**CGGCTAAA**TGAGAGTCGCCGGG
xyrA-X2	ATGTCCCTGTCTTG**AGGCTAAA**GTGGCGCTCAACAA
*xdh*A-A2	AATGAAAAGAACCAG**CGGTTAAA**TTCCATCTGCTTTAG
xynF1-X1	CGGGGTATT**AGGCTAAA**CGTGGCTATCATGGTG
xynF1-X2	TATCATGGTGAAAC**AGGCTGAA**GAACAGTTCCTGGC
xynG2-X1	ATAAAACTGTGCTT**AGGCTAAA**CAACAATTAACTAA
xynG2-X2	CACAGCTCAGACAT**AGGCTAAA**TGGGATCACCACAG
xynG2-X3	TGCTGGGGATAGTC**CGGCTGAA**TTATGGCATATGTA



X1 X2 A2 X1 X2 X1 X2 X3

xyrA *xdhA* *xynF1* *xynG2*

b

+1 +2 +3 +4 +5 +6 +7 +8

cggggtATTAGGCTAAAcgtgg

+1	+2	+3	+4	+5	+6	+7	+8
WT C G T	C A T	C A T	WT G A T	C G A	C G T	WT C G T	C G T