

## Supplementary information

Non-canonical integration events in *Pichia pastoris* encountered during standard transformation analysed with genome sequencing

Jan-Philipp Schwarzhans<sup>1,2</sup>, Daniel Wibberg<sup>3</sup>, Anika Winkler<sup>2</sup>, Tobias Luttermann<sup>1</sup>, Jörn Kalinowski<sup>2</sup>, Karl Friehs<sup>1,\*</sup>

<sup>1</sup> Fermentation Engineering, Bielefeld University, Universitätsstr. 25, Bielefeld, 33615, Germany

<sup>2</sup> Microbial Genomics and Biotechnology, Center for Biotechnology (CeBiTec), Bielefeld University, Universitätsstr. 27, Bielefeld, 33615, Germany

<sup>3</sup> Genome Research of Industrial Microorganisms, CeBiTec, Bielefeld University, Universitätsstr. 27, Bielefeld, 33615, Germany

\*Corresponding author at: [karl.friehs@uni-bielefeld.de](mailto:karl.friehs@uni-bielefeld.de)

|               |     |                                                                                                        |     |
|---------------|-----|--------------------------------------------------------------------------------------------------------|-----|
| MXR1          | 1   | MKGYIFLLAPSSSKASGF <del>AFMGS</del> FRNFFKRSSSHMPVNSSLVSSNIRKSPQDKVVT <u>TVAGGCFWGLEHIYKMHFKDRIVDT</u> | 80  |
| JPS496 Mutant | 1   | MKGYIFLLAPSSSR <del>SNIQRRKVE</del> *-----                                                             | 24  |
| MXR1          | 81  | <u>QVG</u> FANGNVANPTYTQICLGMTHHAEVLQISYNPDVTSFNELIDFFFLVHDPTQKDRQGNDIGSQYRS <del>AVFYLDEKEKEI</del>   | 160 |
| JPS496 Mutant |     | -----                                                                                                  |     |
| MXR1          | 161 | <u>TEQSLAEAQKKWFPHHKIVTQVEKLTSYLDADDYHQDYLMKNPN</u> GYHCPHVLRLKEPKSMSI*                                | 223 |
| JPS496 Mutant |     | -----                                                                                                  |     |

Figure S1: Alignment of the AA sequence of the native *mxr1* protein (GenBank: CCA41165.1) from *P. pastoris* CBS 7435 and the mutant found in strain JPS496, based on the coding sequence found during genome sequencing. Mutations are highlighted in red. The alignment was performed using the Constraint-based Multiple Alignment tool (Cobalt) of the NCBI. Via the NCBI Conserved Domain Database a peptide methionine sulfoxide reductase domain (underlined) was detected in the region AA 56-203 of the native protein.

|                  |     |                                                                                    |     |
|------------------|-----|------------------------------------------------------------------------------------|-----|
| PP7435_CHR2-1130 | 1   | MRVLIHEKTRLIALVSSTHCLLIKIYIKDKNQCAVDFVEKESFSPKHKPLSRNKAHGFLGLINVNEDVFLCVITGKTEVA   | 80  |
| JPS014 Mutant    | 1   | MRVLIHEKTRLIALVSSTHCLLIKIYIKDKNQCAVDFVEKESFSPKHKPLSRNKAHGFLGLINVNEDVFLCVITGKTEVA   | 80  |
| PP7435_CHR2-1130 | 81  | TPTDGETVNKIYNVEFHCLNRDAWDFLELDGNGYPITTSDVNEEALRSRGSTEASDNHLSYEQNPTYELRRLLSNGSFYY   | 160 |
| JPS014 Mutant    | 81  | TPTDGETVNKIYNVEFHCLNRDAWDFLELDGNGYPITTSDVNEEALRSRGSTEASDNHLSYEQNPTYELRRLLSNGSFYY   | 160 |
| PP7435_CHR2-1130 | 161 | STNFDLTSTLQSRDVNSDSLSDSFHLDYMWNSYMMKEVVNFRDRLPTDSKKILDRNGFLT TVIRGFAETFRTRIGHQKC   | 240 |
| JPS014 Mutant    | 161 | STNFDLTSTLQSRDVNSDSLSDSFHLDYMWNSYMMKEVVNFRDRLPTDSKKILDRNGFLT TVIRGFAETFRTRIGHQKC   | 240 |
| PP7435_CHR2-1130 | 241 | NATIISKQSWKRAGTRYNARGIDDEGYVANFVETELILHSDKDFIYAYTEVRGSVPIFWEQDTALVNPKVITITRSLEATEP | 320 |
| JPS014 Mutant    | 241 | NATIISKQSWKRAGTRYNARGIDDEGYVANFVETELILHSDKDFIYAYTEVRGSVPIFWEQDTALVNPKVITITRSLEATEP | 320 |
| PP7435_CHR2-1130 | 321 | VFEKHFAALNGKYGPVHIVNLLSTKPSEIGLSNTYRKHF EIVNKKGSPQAYLTEFDFHKETGKNYALATKVIPFLEES IY | 400 |
| JPS014 Mutant    | 321 | VFEKHFAALNGKYGPVHIVNLLSTKPSEIGLSNTYRKHF EIVNKKGSPQAYLTEFDFHKETGKNYALATKVIPFLEES IY | 400 |
| PP7435_CHR2-1130 | 401 | DFDYFSYDVKNQKVLTLQKGVFRTNCLDCLDRTNVVQQVISNATLNMFLQRHNLNTNYDSDLFNKHNTLWADHGDAISQI   | 480 |
| JPS014 Mutant    | 401 | DFDYFSYDVKNQKVLTLQKGVFRTNCLDCLDRTNVVQQVISNATLNMFLQRHNLNTNYDSDLFNKHNTLWADHGDAISQI   | 480 |
| PP7435_CHR2-1130 | 481 | YTGTNALKSSFTRSGKMGIAGALSDATKSISRMYINNFDKAKQVTIDTLLGKMSNQVEVRIFDPVSDHVDNELSKLKSQ    | 560 |
| JPS014 Mutant    | 481 | YTGTNALKSSFTRSGKMGIAGALSDATKSISRMYINNFDKAKQVTIDTLLGKMSNQVEVRIFDPVSDHVDNELSKLKSQ    | 560 |
| PP7435_CHR2-1130 | 561 | FSAEDDIRIFTGSYNLGGTAYADDFTDWLFPKENG GIEGAPDVVILGFQEVVELTASNILNSDSSRS HYWSEEIKTQLNK | 640 |
| JPS014 Mutant    | 561 | FSAEDDIRIFTGSYNLGGTAYADDFTDWLFPKENG GIEGAPDVVILGFQEVVELTASNILNSDSSRS HYWSEEIKTQLNK | 640 |
| PP7435_CHR2-1130 | 641 | ISSSKYILLRSEQMTSLLLFFIKEDKMPKV TQVEGCSKKTGLGGITANKGAVALRFSFGSTTFCLLNSHLAAGLNSVVE   | 720 |
| JPS014 Mutant    | 641 | ISSSKYILLRSEQMTSLLLFFIKEDKMPKV TQVEGCSKKTGLGGITANKGAVALRFSFGSTTFCLLNSHLAAGLNSVVE   | 720 |
| PP7435_CHR2-1130 | 721 | RNNDFTTISQGIRFSRNKTIYDHDCVIWLGD LNYRVPLPNELVRSSALNGVYDELLAEDQLKTEMVHKGAFADFYEMKIN  | 800 |
| JPS014 Mutant    | 721 | RNNDFTTISQGIRFSRNKTIYDHDCVIWLGD LNYRVPLPNELVRSSALNGVYDELLAEDQLKHPKTKG*-----        | 789 |
| PP7435_CHR2-1130 | 801 | FLPTYKYDKGTSVFDTSEKQRPVSWTDRILYRGKRLQQVNYNSVQSITISDHPKIYGTGFAHVTVYDERTKLTLMRKIYD   | 880 |
| JPS014 Mutant    |     | -----                                                                              |     |

|                  |      |                                                                                  |      |
|------------------|------|----------------------------------------------------------------------------------|------|
| PP7435_CHR2-1130 | 881  | DYRRDPKSEASSASSNGDKFGDLIDFNDASSSSNTSIYTETEGKQVSQPVPGHFPPPPPPPRNRVTPPQPSTPETPPPPG | 960  |
| JPS014 Mutant    |      | -----                                                                            |      |
| PP7435_CHR2-1130 | 961  | YALNMAPLIPSRANSNNTRPSSPFVSTDRAKAEISTKPVRPMVPDKPSSLKAMSPENTIQTVPPVQKAISTDSLNTAS   | 1040 |
| JPS014 Mutant    |      | -----                                                                            |      |
| PP7435_CHR2-1130 | 1041 | AAPPPPPPRKQESINSSMSAFPVLMPPKKK*                                                  | 1070 |
| JPS014 Mutant    |      | -----                                                                            |      |

Figure S2: Alignment of the AA sequence of the native PP7435\_Chr2-1130 protein (GenBank CCA38807.1) from *P. pastoris* CBS 7435 and the mutant found in strain JPS014, based on the coding sequence found during genome sequencing. Mutations are highlighted in red, with the silent mutation c.G2346A in green. The alignment was performed via Cobalt of the NCBI. Two domains (underlined) were predicted in the native protein, using the NCBI Conserved Domain Database: a *SacI* homology domain (AA 56-381) and a catalytic inositol polyphosphate 5-phosphatase (INPP5c) domain (AA 567-861).

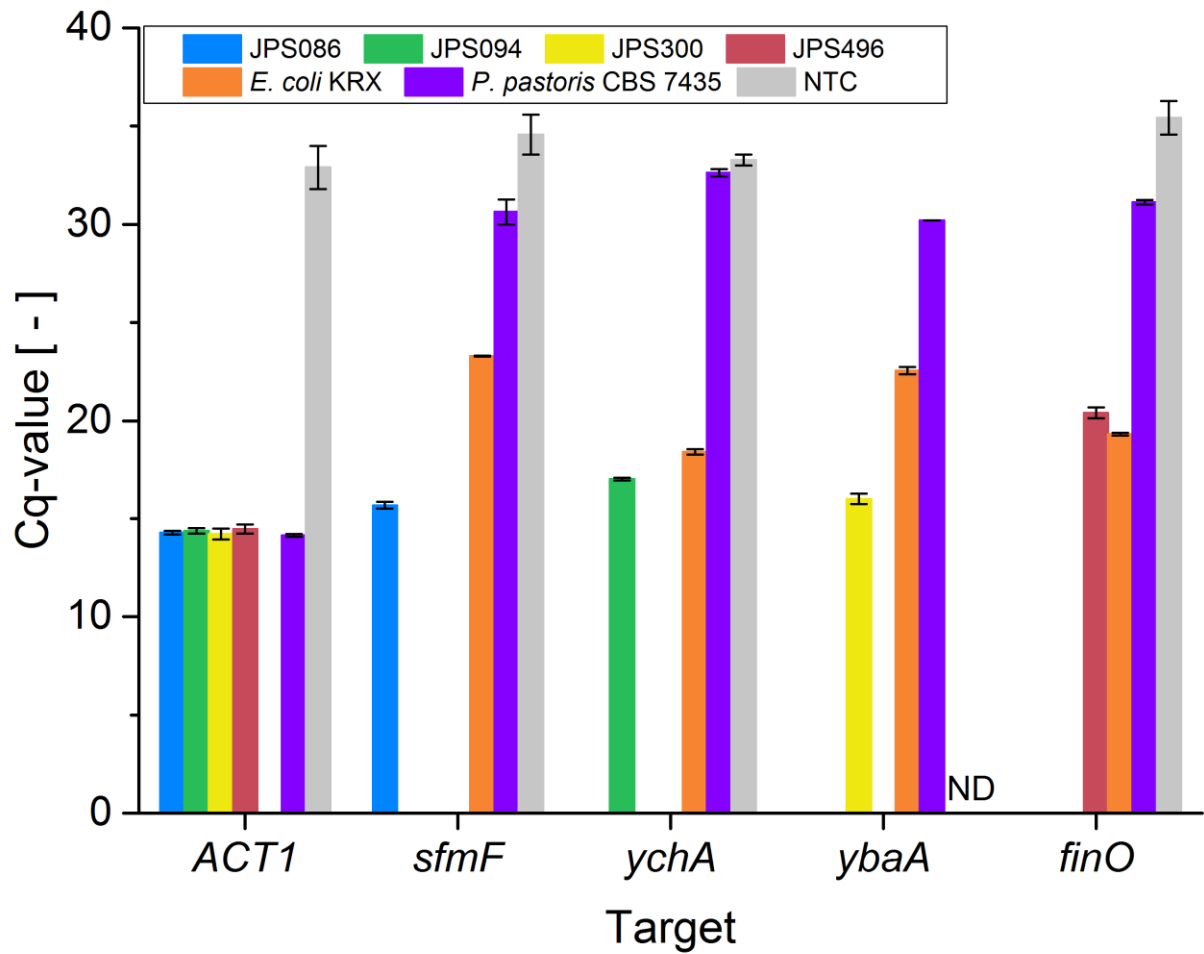


Figure S3: Cq-values measured during qRT-PCR experiments for the different target amplicons in *E. coli* KRX, *P. pastoris* CBS 7435 and the mutant strains JPS086, JPS094, JPS300 and JPS496. Additionally, the no template control (NTC) is shown. The targets are the housekeeping gene *ACT1* of *P. pastoris* as well as the *E. coli* genes *sfmF* (fimbrial protein), *ychA* (adhesin *AidA* precursor), *ybaA* (signaling protein) and *finO* (fertility inhibition). For the target *ybaA* no detectable (ND) signal was found in the NTC after 40 cycles. The melting curve analysis after 40 cycles revealed that in the *P. pastoris* CBS 7435 samples secondary amplicons were the cause for Cq-values < 33.3. Error bars indicate the standard deviation with n = 3.