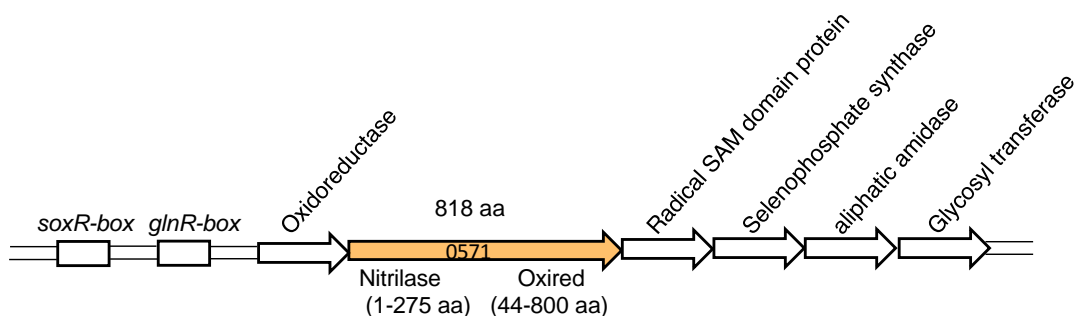# A

B

MTTLAAVSANFTRDLEQNFALIADLAEQAREKGVDFLVLPEAAIGGYLSSLG
NHGDTVRATSRSLPPAIRLDGPEIARVQQIVGDLVVAIGFCELADDGETRYN
AAAVLDRFDVYGSYRKVHQPLGEGMSYSAGSDYGVFDTPVGRVGLQICYDK
AFPEAARLMALGGAQIIASLSAWPAARTATAQNLQDDRWTYRFNLFDMARA
LDNQVFWVASNQSGSFGSLRYVGNAKVVDPGGNVLATTLLDSGMAVAEVDI
DETFRTMRAGMFHLRDRRPDVYAPLTDSDGPRTAKWRELAHA**U**DDLRSPV
AGRQHPALLFPEPGDPRPSDRGNALRRRRFRGALEPSTGRGERAGTRQVR
LRLHLGGRDPSAHRQRRPAVQRGCGARGRRHAPRTGAVMTTHVPVAVIGG
GQAGLSVSWYLVRAGIEHIVIESKTPMHAWADTRWDNFTLVTPNWHCRLPG
YPYAGPDPDGFMTRDEVVDWLAGWLDTFDPPLRNHTQVTRLQNRAGGGFE
VTLRDESGQSTLTCDHAVIATGGYPVPVIPSYAAQLDETILQIHSEQYRNAGT
LPDGAVLVVGTGQSGAQIAEDLHLAGRRVHLAVGGAPRVARFYRGRDCMT
WLADMGVYDRPAQQYPGGQAAIEKTNHYVTGRDGGRDVDLRQFATEGMRL
YGTLADGKDSTLRFEPTLAEALDHADSVYNSICSDIDAHIERNGIDAPPASRY
EPVWKPETETTTLDLAAAGITSIVWAIGYRPDYRWIAASAFDGAGRPMQTRG
ITNVAGLSFVGLPWMHTWGSGRFLGIDRDARHIAATIISSYHESVLRLAMKV

Figure S5: Possibility of MSMEG_0571 in *gpr* locus as a new selenoprotein. **A.** Schematic representation of new ORFs due to incorporation of selenocysteine at the UGA codon of MSMEG_0571, resulting in a single polypeptide of 818 amino acids, terminating at the UAG codon of original MSMEG_0569 ORF. BLAST search of this protein identifies a nitrilase and flavin-dependent oxidoreductase (Oxired) domains at the N- and C- terminals, respectively. The position of amino acids encompassing the domains are also indicated. **B.** Sequence of the possible selenoprotein encoded by *gpr*. Seleneocysteine (U) is indicated in red, while the segments of the protein that were originally parts of MSMEG_0571 (blue) and MSMEG_0569 (brown) are color coded. A linker (black) between the two is out of frame with the originally annotated MSMEG_0570.