

Supplementary Information

Conserved molecular structure of the centromeric histone CENH3 in *Secale* and its phylogenetic relationships

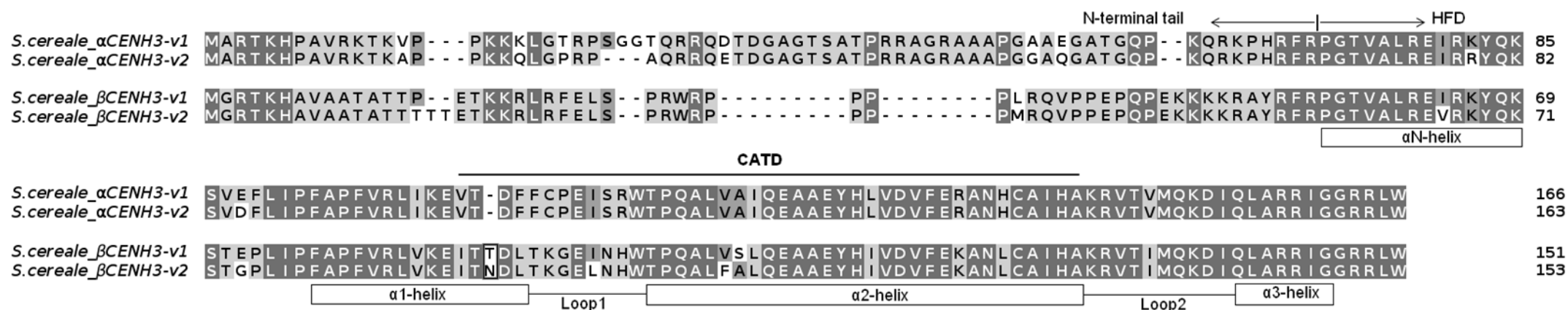
E.V.Evtushenko¹, E.A.Elisafenko², S.S.Gatzkaya¹, Y.A.Lipikhina¹, A.Houben³, A.V.Vershinin^{1*}

¹ Institute of Molecular and Cellular Biology SB RAS, Novosibirsk, 630090, Russia

² Institute of Cytology and Genetics SB RAS, Novosibirsk, 630090, Russia

³ Leibniz Institute of Plant Genetics and Crop Plant Research (IPK), Gatersleben, 06466 Stadt Seeland, Germany

* corresponding author: avershin@mcb.nsc.ru



Supplementary Fig. S1. Multiple alignment of the amino acid sequences of CENH3 variants in *S. cereale* (variety “Imperial”)

The amino acids that occur in the β CENH3 HFD, but not in the α CENH3, are framed.

Supplementary Table S1. Nonsynonymous (K_a) to synonymous (K_s) nucleotide substitutions in the NTT domain of βCENH3s of *Secale* subspecies

	CENH3 <i>S. cereale</i>	CENH3 <i>S.afghanicus</i>	CENH3 <i>S.vavilovii</i>	CENH3 <i>S. strictum</i>	CENH3 <i>S.kuprijanov</i>	CENH3 <i>S.africanum</i>
CENH3 <i>S.sylvestre</i>	0.022/0.048 0.457	0.018/0.027 0.668	0.020/0.048 0.421	0.017/0.029 0.579	0/0.020/0.048 0.420	0.017/0.026 0.660
CENH3 <i>S.cereale</i>		0.020/0.062 0.322	0.0009/0.0064 1.48	0.019/0.066 0.291	0.0092/0.009 0.984	0.019/0.064 0.309
CENH3 <i>S.afghanicus</i>			0.012/0.063 0.296	0.014/0.021 0.537	0.019/0.063 0.296	0.010/0.017 0.581
CENH3 <i>S.vavilovii</i>				0.017/0.067 0.266	0.007/0.003 2.07	0.018/0.066 0.279
CENH3 <i>S.strictum</i>					0.018/0.067 0.265	0.008/0.017 0.515
CENH3 <i>S.kuprijanovii</i>						0.018/0.066 0.278

Bold numbers indicate the ratio $\omega = K_a : K_s > 1$

Supplementary Table S2. Description of amino acid substitutions in CENH3 sequences

#	CENH3 domain	Codon	Species, subspecies	Codon substitution	Amino acid substitution	Substitution Character	P-value	Significance level	Method
1*	α CENH3 v-1, NTT	34*	<i>S. ancesrale</i> (AO) <i>S. dighoricum</i> (AO) <i>S. anatolicum</i> (PO)	ACA→GCA ACA→CCA ACA→CCA	T→A T→P T→P	non-synonymous	0.0433	0.05	MEME
2*	α CENH3 v-1, NTT	22	<i>S. cereale</i> (AO, AS)	ACG→ACC	no change (T)	synonymous	0.0123	0.1	SLAC
3*	α CENH3 v-1, NTT	48	<i>S. strictum</i> (PO, PS)	CGT→CGG	no change (R)	synonymous	0.0557	0.1	SLAC
4*	α CENH3 v-1, HFD	92	<i>S. cereale</i> (AO, AS)	CCT→CCC(G)	no change (P)	synonymous	0.0370	0.1	SLAC
5*	α CENH3 v-1, HFD	136	<i>S. strictum</i> (PO, PS)	AGG→CGG	no change (R)	synonymous	0.0017	0.1	SLAC
6*	β CENH3 v-1, NTT	51	<i>S. strictum</i> (PO) <i>S. afghanicum</i> (AO)	GCG→GCT GCG→GCT	no change (A) no change (A)	synonymous	0.0370	0.1	SLAC
7	α CENH3 v-1, NTT	10	<i>S. cereale</i> (AO) (10)** <i>S. strictum</i> (PO) (33) <i>S. anatolicum</i> (PO) (40) <i>S. sylvestre</i> (AS) (14)	AGG→GGG	R→G	non-synonymous	—	—	—
8	α CENH3 v-1, NTT	31	<i>S. strictum</i> (PO) (11) <i>S. anatolicum</i> (PO) (5) <i>S. sylvestre</i> (AS) (7)	CGG→GGG	R→G	non-synonymous	—	—	—
9	α CENH3 v-1, NTT	33	<i>S. cereale</i> (AO) (22) <i>S. strictum</i> (14) <i>S. sylvestre</i> (AS) (7)	GAC→GAG	D→E	non-synonymous	—	—	—
10	α CENH3 v-1, HFD	104	<i>S. strictum</i> (PO) (20) <i>S. anatolicum</i> (PO) (10) <i>S. africanum</i> (PS) (10)	ACC→GCC	T→A	non-synonymous	—	—	—
11	α CENH3 v-1, HFD	110	<i>S. strictum</i> (PO) (18) <i>S. anatolicum</i> (PO) (11) <i>S. africanum</i> (PS) (5)	GAA→AAA	E→K	non-synonymous	—	—	—
12	α CENH3 v-1, HFD	130	<i>S. strictum</i> (PO) (12) <i>S. anatolicum</i> (PO) (11) <i>S. africanum</i> (PO) (14)	CTC→CAC	L→H	non-synonymous	—	—	—
13	β CENH3 v-1, NTT	17	<i>S. sylvestre</i> (AS) (80)	AAG→ACG	K→T	non-synonymous	—	—	—
14	β CENH3 v-1, NTT	41	<i>S. sylvestre</i> (AS) (67)	GAG→CAG	E→Q	non-synonymous	—	—	—

* CENH3 sequences are under selective pressure

** relative number of substitution (%)

Supplementary Table S3. List of primers used for amplification of full-length sequences and separate domains of CENH3

Primer number	Primer Sequence, 5' – 3'	Amplified sequence of <i>CENH3</i>
1	ATGGCCCGCACCAAGCAC	Coding part of the rye α <i>CENH3</i> gene
2	GCATCACCAAAGCCTCC	
3	TGGGTCGCACGAAGCAC	Coding part of the rye β <i>CENH3</i> gene
4	TCACCAAAGCCTTCTCCCC	
5	GAAACTCGACCGACTTCTG	NTT of the rye α <i>CENH3</i> gene (with primer #1)
6	ATGGGTCGCACGAAGCAC	NTT of the rye β <i>CENH3</i> gene
7	CGTGAGGTCTGGTGGTGATTT	
8	GTGGCACTGCGGGAGATCAGGA	HFD of the rye α <i>CENH3</i> gene (with primer #2)
9	GCGCTGCGGGAGATCAG	HFD of the rye β <i>CENH3</i> gene (with primer #7)
10	GTRGCRCTGCGGGAGATCAGGA	HFD of the <i>CENH3</i> gene from <i>Triticum</i> and <i>Aegilops</i> species
11	CTBGCRAGYTGYATGTCCTTTT	