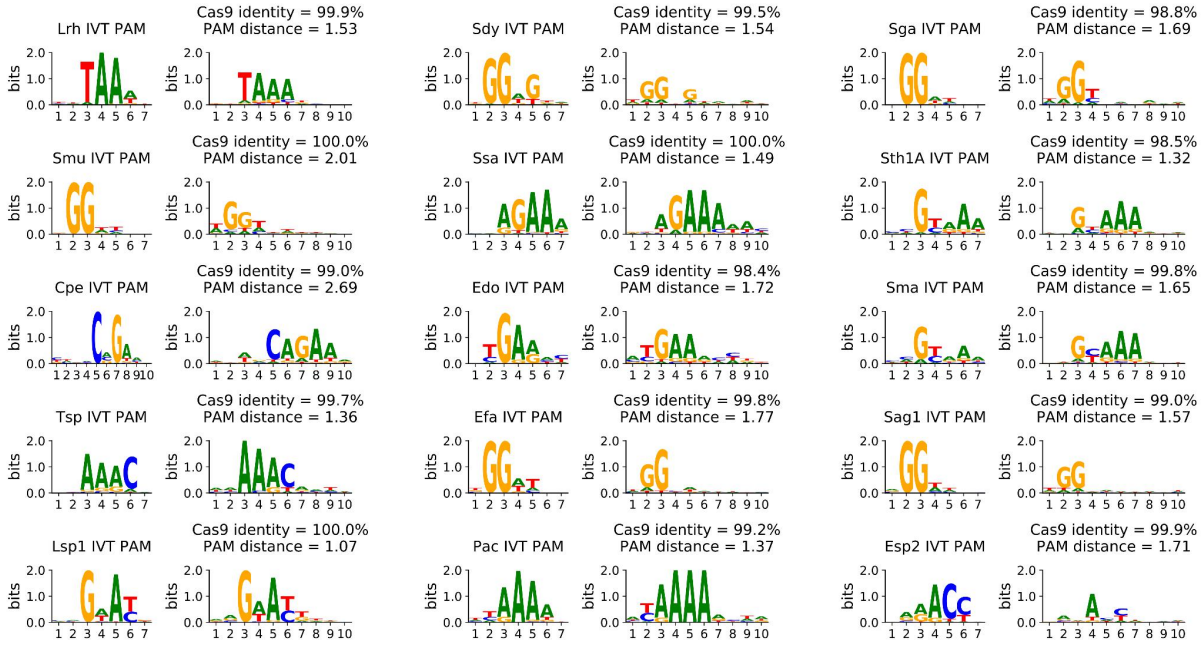
Supplementary Information

Tailored identification of Cas9 proteins with specific PAM requirements using a prediction computational pipeline

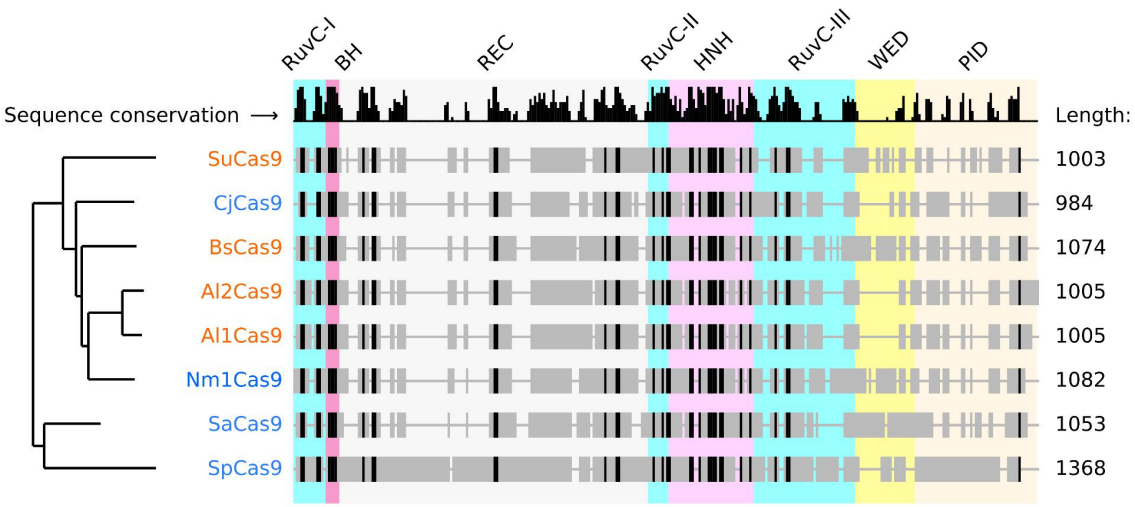
Table of contents

Supplementary Figure 1



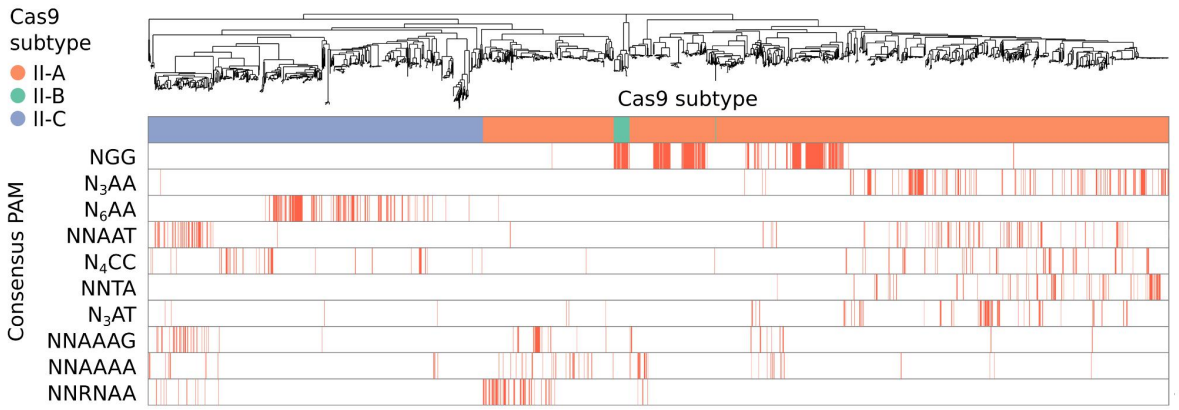
Supplementary Figure 1: Predicted and experimentally identified PAM logos analyzed for selected Cas9s reported by Gasiunas *et al.* [13]. For each indicated Cas9 the *in vitro* (IVT) experimentally derived (left) and *in silico predicted* (right) logos are reported. Percentages of sequence identity and measure of the distance between predicted and true PAMs (see **Methods** in main text) are reported for each Cas9.

Supplementary Figure 2



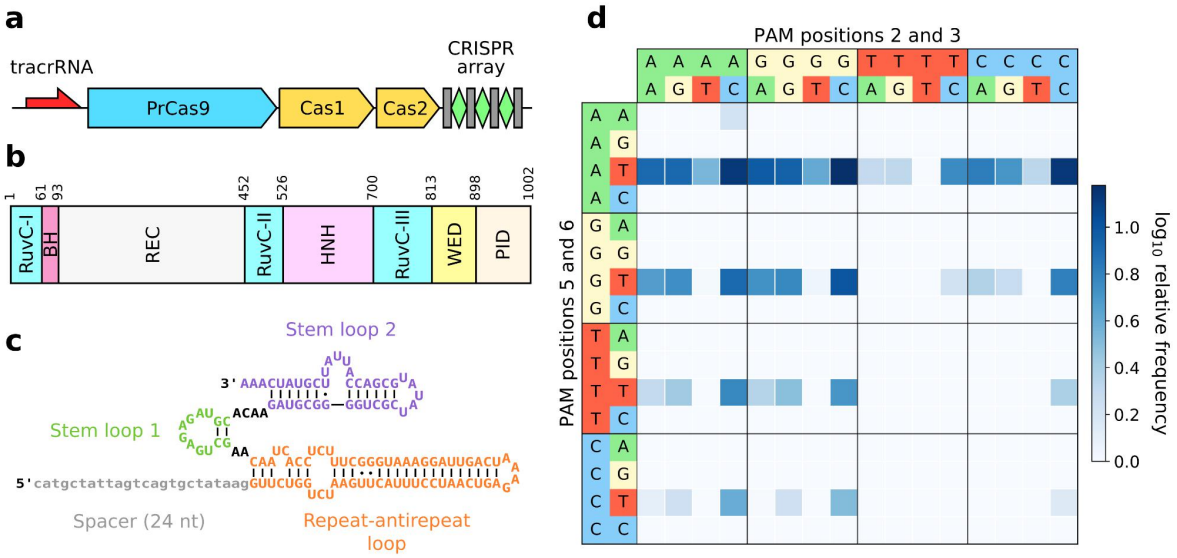
Supplementary Figure 2: Protein sequence alignments of newly discovered and described Cas9s. a) Phylogenetic tree of selected Cas9 proteins mostly used for genome editing applications (blue) and newly characterized Cas9 proteins (orange). Protein alignments: grey aligned protein sequences, black conserved sequences, colored conserved domains. Length: number of amino acids.

Supplementary Figure 3



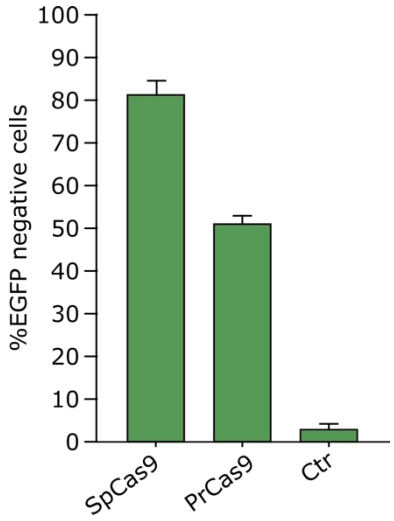
Supplementary Figure 3: Associations between the 10 most abundant PAM clusters and the Cas9 phylogenetic tree.

Supplementary Figure 4



Supplementary Figure 4: PrCas9 identified for using as PAM the P23H-RHO mutation. a) Organization of the PrCas9 CRISPR-Cas locus. b) Predicted protein domain organization of PrCas9. c) Structure of the single guide RNA (sgRNA) of PrCas9. d) PAM heatmap for PrCas9, showing the nucleotide preference for positions 2, 3, 5 and 6 (no nucleotide preference is present at other positions). The preferred PAM is NRVNRT, V = A, C or G; R = G or A.

Supplementary Figure 5



Supplementary Figure 5: Editing activity of PrCas9 in mammalian cells. The editing activity of PrCas9 was evaluated by using an EGFP disruption assay by measuring the fluorescence of U2OS cells stably expressing EGFP and transfected with SpCas9, PrCas9 or with control plasmids (Ctr).

SUPPLEMENTARY METHODS

EGFP disruption assay

The assay to measure PrCas9 editing activity was performed using the U2OS.EGFP cells (a kind gift of Claudio Mussolino, University of Freiburg) carrying a single integrated copy of an EGFP reporter gene. U2OS.EGFP cells were cultured in DMEM (Life Technologies) supplemented with 10% FBS (Life Technologies), 2 mM GlutaMax (Life Technologies) and penicillin/streptomycin (Life Technologies) at 37°C and 5% CO2 in a humidified atmosphere. Cells tested mycoplasma negative (PlasmoTest, Invivogen). To perform the experiments, 200,000 U2OS.EGFP cells were nucleofected with 1 ug of pX-PrCas9 plasmid bearing a gRNA designed to target EGFP using the 4D-Nucleofector™ X Kit (Lonza), DN100 program, according to the manufacturer's protocol. After electroporation, cells were plated in a 96-well plate. After 48 hours cells were expanded in a 24-well plate. EGFP knock-out was analyzed 4 days after nucleofection using a BD FACSCanto (BD) flow cytometer.

Supplementary Table 1. Description of metagenomic datasets from which 54,169 additional MAGs were retrieved

Study ID	PMID / DOI	Environment	Number of samples	Number of MAGs
PaoliL_2021	10.1101/2021.03.24.436479	Ocean	6261	32366
BeresfordJonesBS_2021	34971560	Mice gut	1043	18733
HeroldM_2020	33077707	Wastewater	787	2098
AlbaneseD_2021	33741058	Antarctic desert	33	562
LeechJ_2020	33172966	Fermented food	70	228

LandisEA_2021	33496265	Bakery	43	107
WylensekD_2020	33319778	Pig gut	38	75
KastmanEK_2016	27795388	Dairy	19	56
PfeferT_xxxx	-	Dairy	35	54
LiZ_2018	30166168	Fermented foods	11	49
ArikanM_2020	31957879	Fermented foods	12	35
PatroJN_2016	27303722	Probiotics	10	33
ZhaoCC_2020	32247457	Fermented foods	5	32
ChaconVargasK_2020	32934253	Fermented foods	8	32
PothakosV_2020	10.1016/j.crbiot.2020.02.001	Fermented foods	10	29
PasolliE_2020	32451391	Dairy;Probiotics	17	28
KumarJ_2019	31428064	Fermented foods	4	21
VerceM_2019	30918501	Fermented foods	6	21
YulandiA_2020	33308279	Fermented foods	2	20
EscobarZepedaA_2016	27052710	Dairy	1	18
SulaimanJ_2014	25400624	Brine	7	18
DuruIC_2018	29803134	Dairy	6	17
Ferrocinol_2018	29196291	Fermented foods	11	17
DuR_2020	32276974	Alcohol	3	16
LiZ_2019	31500701	Fermented foods	6	13
PorcellatoD_2016	10.1016/j.idairyj.2016.05.005	Dairy	12	13
EinsonEJ_2018	30171008	Fermented foods;Fruit and Vegetables	4	12
LordanR_2019	10.1016/j.jff.2019.01.029	Dairy	5	11

SalvettiE_2016	27445999	Fruit and Vegetables	2	11
LeonardSR_2016	27930729	Fruit and Vegetables	7	9
DeRoosJ_2020	32765478	Alcohol	4	8
YasirM_2020	33233218	Dairy	2	3
SomervilleV_2019	31238873	Dairy	1	2
CrovadoreJ_2017	28572315	Other	2	2

142

143

Supplementary Table 2. Sequences of the primers used for NGS library preparation in the *in vitro* PAM assay	
Primer name	Sequence (5' → 3')
F4a	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGCTGCTGAACCGCTCTTCCGATC
F4b	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGTAAGACTGCTGAACCGCTCTTCCGATC
F4c	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGGCTAGACCTAATGTGATCTGCTGAACCGCTCTTCCGATC
R3	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGTCTGCGTTCTGATTTAATCTGTATCAGGC

144

145

146

Supplementary Table 3. Sequences of the oligonucleotides used for spacers cloning in the expression plamids.		
Plasmid	**Oligonucleotide 1 (5'-3')**	**Oligonucleotide 2 (5'-3')**
pX-SpCas9-sgRNA-GFPB	CACCGGGCACGGGCAGCTTGCCGG	GAACCCGGCAAGCTGCCCGTGCCC

pX-PrCas9-sgRNA-GFP	CACCGCCCGAAGGCTACGTCCAGGAGCG	GAACCGCTCCTGGACGTAGCCTTCGGGC
pX-PrCas9-sgRNA-RHO-P23H	CACCGCAGCCAGGTAGTACTGTGGGTACT	GAACAGTACCCACAGTACTACCTGGCTGC

147
148
149

Supplementary Table 4. PrCas9 amino acid sequence.

MKMQDSVSKMKYRLGIDLGTTSLGWAMLRLDEQNEPYAVIRAGVRIFNNGRDPKTEASL
AVARRLARQQRRTRDRKIRRKERLIGELVDMGFFPKDPVKRRQLASLDPFKLRTEALDRA
LSPEEFARAIFHLARRRGFKSNRKTDSGDTESSKMKEAIKRTLNELQNKGFRTVGEWLN
MRHQQRLGTRSRIKNVPTGSGKQTTAYDFYLNRFMIEYEFDRIWEKQSQMNPGLFTNER
KAILKDIIFYQRPLRPVEPGRCTFMPDNPRAPLALPQQQDFRIYQEVNNLRKIDPTSLLEVN
LTLPERDRIVELLQRKPALTFDAVRKALCFNGTFNLEGENRSELKGNLTNCALAKKKLFGE
SWYSFDAHKRFEIVEHLLQEESEENLVSWLQKECNLSEEYAKNVASVRLPAGYGALCQE
ALDLILPYLKAEVITYDKAVQKAGMNHSELTLAQETGEILPELPYYGQYLKRHVGFGTGKP
EDSAEKRYGKIPNPTVHIALNQLRTVVNALIRRYGKPTQIVIELARELKQNKKAKDQYRIEM
NHNQNRNERIRADISMILGINPENVKRKDIEKQILWEELNLKDATARCCPYSGKQISAEMLF
TDEVEIDHILPFSRTLDDSKNNKVVCIREANRIKGNRTPWEARKDFEKRGWSVEAMTARA
QAMPKAKRFRFAEDGYKVWLKDFDGFEARALTDTQYMSRVAREYLQLICPGQTWSVPG
QLTGMLRRFLGLNDILGVNGEKNRDDHRHHAVDACVIALTDRSMLQRISTASARAENKHL
TRLLESFPAPWATFYEHVTRAVKSICVSHKPEHAYQGAMNEQTAYGLRPDGYVKYRQNG
KVEHKKLNVIPQVSVKGTWRHGLNSDGSLKAYKGLKGGSNFCIEIVMGEGGRWEGDVIT
TYEAYQIVRAKGEAALYGSVSRSGKPLVMRLMQKDIVEMTLADGRCKMLLYIITQNKQMF
FYRIENAGGGREDVSKRPGSLQKALAKKIIVSPIGDFRKEKL*

150
151
152
153
154

155