Supplementary material for the paper "Evolution of increased complexity in a molecular machine" Finnigan, et al. Nature 2011

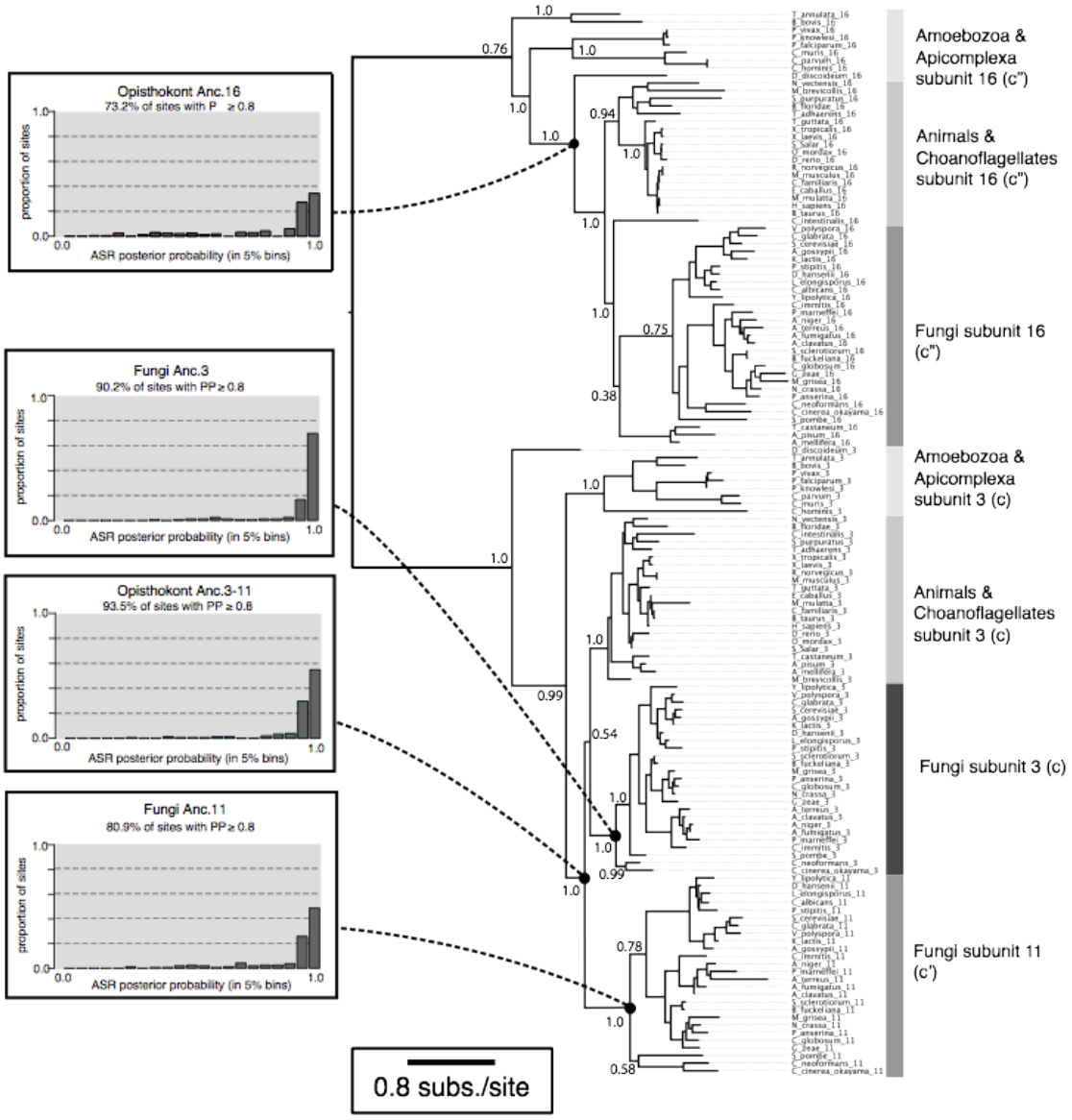
Supplement S1

GenBank accession IDs for protein sequences used in this study. The IDs are labeled with the first letter of their genus, the full name of their species, and an integer number. 3, 11, and 16 indicate homology to yeast subunits c, c', and c", respectively.

M_musculus_16 NP_291095
O_mordax_3 ACO09611
D_rerio_3 NP_001098606
M_grisea_16 XP_369356
M_grisea_11 XP_366989
A_niger_3 XP_001399935
C_glabrata_3 XP_447321
A_terreus_11 XP_001214955
A_terreus_16 XP_001211600
C_parvum_16 XP_627363
P_vivax_16 XP_001616329
T_castaneum_3 XP_967959
X_tropicalis_3 NP_988893
G_zeae_3 XP_390178
L_elongisporus_3 XP_001526092
A_fumigatus_3 XP_001263225
M_grisea_3 XP_365764
S_pombe_3 NP_594799
C_albicans_3 XP_721376
B_bovis_16 XP_001612047
G_zeae_11 XP_388749
C_muris_3 XP_002141961
S_purpuratus_3 XP_797801
C_immitis_16 XP_001246494
M_mulatta_16 XP_001097275
Y_lipolytica_3 XP_505831
B_fuckeliana_16 XP_001552198
C_immitis_11 XP_001242880
A_mellifera_16 XP_392599
L_elongisporus_11 XP_001523616
T_annulata_16 XP_953463
N_vectensis_3 XP_001637733
S_Salar_16 NP_001134021
L_elongisporus_16 XP_001525467
C_neoformans_16 XP_773114
C_neoformans_11 XP_778255
P_knowlesi_16 XP_002261350
A_terreus_3 XP_001213329
A_pisum_3 NP_001155531
S_Salar_3 NP_001154112
P_marneffei_3 XP_002152865
P_stipitis_3 XP_001387092
A_gossypii_3 NP_984787
C_hominis_3 XP_667190
C_muris_16 XP_002142524
X_laevis_16 NP_001087741
M_mulatta_3 XP_001088617
M_brevicollis_16 XP_001742805
K_lactis_3 XP_454966
T_castaneum_16 XP_975026

C_intestinalis_16 XP_002131348
E_caballus_3 XP_001915231
G_zeae_16 XP_385476
C_globosum_3 XP_001229170
A_clavatus_3 XP_001271234
N_crassa_11 XP_965807
N_crassa_16 XP_964449
X_tropicalis_16 NP_001017064
P_anserina_11 XP_001907168
P_anserina_16 XP_001910317
H_sapiens_16 AAP36886
T_annulata_3 XP_952989
A_gossypii_11 NP_985409
A_gossypii_16 NP_983473
A_pisum_16 NP_001155679
N_crassa_3 XP_961418
V_polyspora_3 XP_001642185
T_adhaerens_3 XP_002112261
S_cerevisiae_3 NP_010887
P_marneffei_16 XP_002145395
D_hansenii_3 XP_460869
S_pombe_11 NP_593600
T_guttata_3 ACH45347
S_pombe_16 NP_594516
B_fuckeliana_3 XP_001553113
T_adhaerens_16 XP_002114348
S_cerevisiae_16 NP_011891
R_norvegicus_3 NP_033859
S_cerevisiae_11 NP_015090
B_taurus_3 NP_001017954
C_familiaris_3 XP_537002
P_anserina_3 XP_001911041
P_marneffei_11 XP_002147471
P_knowlesi_3 XP_002259621
T_guttata_16 NP_001232246
B_fuckeliana_11 CCD51873
C_cinerea_okayama_3 XP_001835649
P_falciparum_16 XP_001350256
C_intestinalis_3 XP_002132074
A_niger_11 XP_001391591
A_niger_16 XP_001397102
C_parvum_3 XP_627909
S_purpuratus_16 XP_790651
B_floridae_3 XP_002598155
Y_lipolytica_16 XP_505205
A_mellifera_3 NP_001011570
A_clavatus_16 XP_001275839
N_vectensis_16 XP_001638230
A_clavatus_11 XP_001274195
S_sclerotiorum_3 XP_001588693
P_falciparum_3 XP_001351750
D_discoideum_16 XP_644318
C_cinerea_okayama_11 XP_001835902
D_hansenii_16 XP_460013
D_hansenii_11 XP_458901
C_cinerea_okayama_16 XP_001830694
P_vivax_3 XP_001613765
C_immitis_3 XP_001239974
A_fumigatus_16 XP_755891

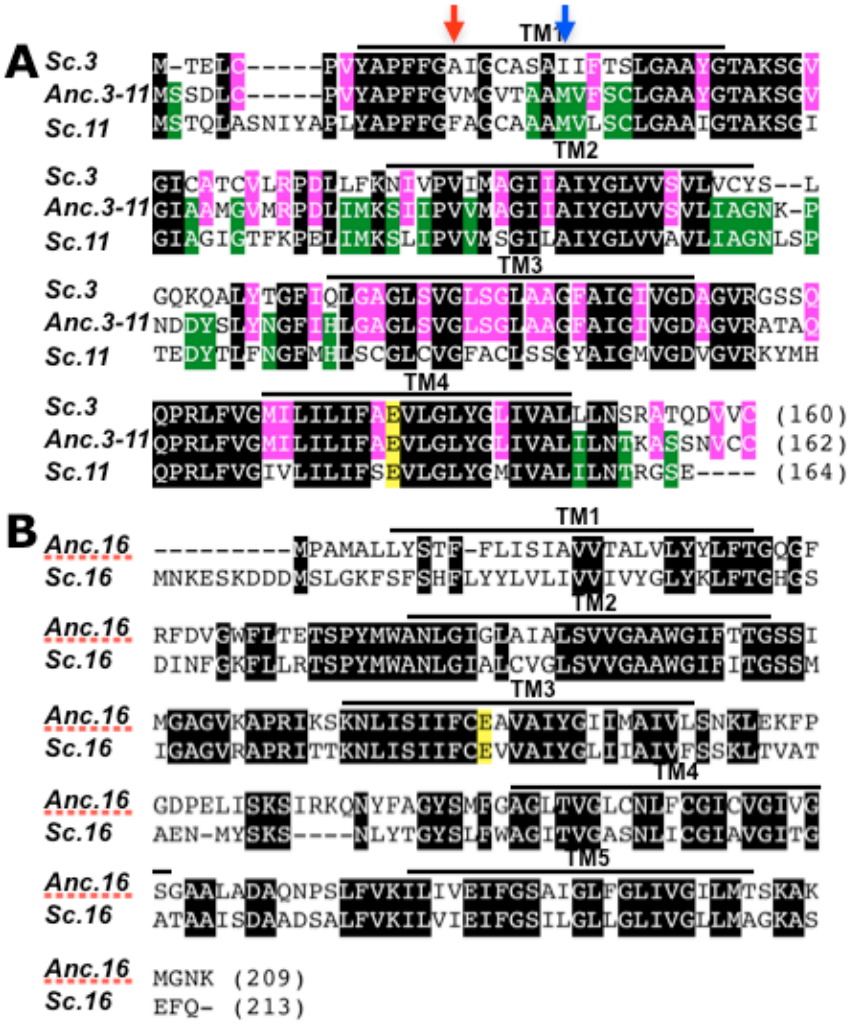
C_neoformans_3 XP_772642
H_sapiens_3 AAP36127
A_fumigatus_11 XP_753781
Y_lipolytica_11 XP_504637
M_brevicollis_3 XP_001743042
D_discoideum_3 XP_644319
P_stipitis_16 XP_001386908
C_albicans_16 XP_722165
C_albicans_11 XP_721376
P_stipitis_11 XP_001382501
X_laevis_3 NP_001082675
C_glabrata_11 XP_445959
E_caballus_16 XP_001916016
C_glabrata_16 XP_447739
V_polyspora_11 XP_001645235
V_polyspora_16 XP_001646358
C_hominis_16 XP_665533
C_globosum_11 XP_001222467
C_globosum_16 XP_001223715
O_mordax_16 ACO10130
D_rerio_16 NP_955855
B_bovis_3 XP_001609797
B_floridae_16 XP_002610356
M_musculus_3 NP_033859
C_familiaris_16 XP_539645
K_lactis_11 XP_452911
K_lactis_16 XP_454470
B_taurus_16 NP_001033127
S_sclerotiorum_16 XP_001590765
S_sclerotiorum_11 XP_001595091
R_norvegicus_16 AAH09169



Supplement S2

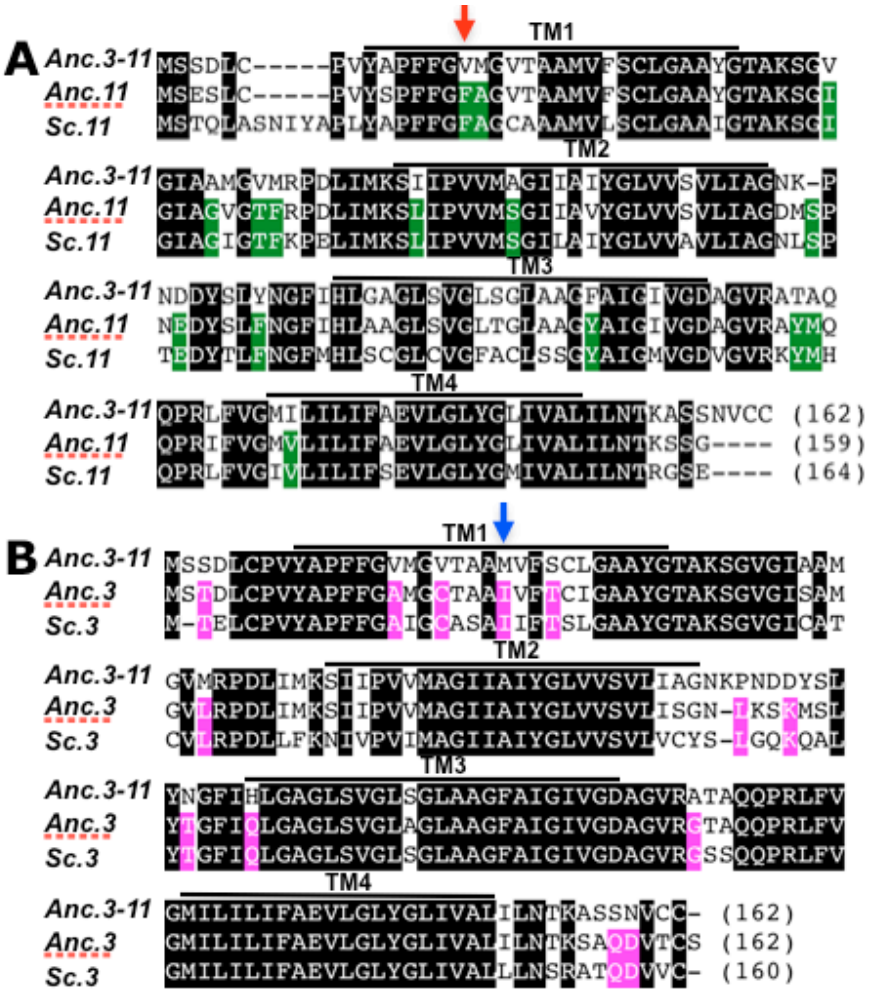
Maximum likelihood phylogeny of protein sequences for V-ATPase subunits 3, 11, and 16. Terminal taxa names correspond to sequences in Supplemental S1. Decimals on internal branches are approximate likelihood ratios with the nonparametric Shimodaira–Hasegawa (SH) correction (Anisimova et al., 2011). The inset bar graphs show support for ancestral reconstructions of ancestors Anc 3-11, Anc.16, Anc.3, and Anc.11.

M. Anisimova, M. Gil, J.-F. Dufayard, C. Dessimoz, and O. Gascuel. Survey of branch support methods demonstrates accuracy, power, and robustness of fast likelihood-based approximation schemes. *Systematic Biology*, 60(5), 2011.



Supplement S3

Protein sequence alignment of contemporary *S. cerevisiae* proteolipids with reconstructed Anc.3-11 and Anc.16. Potential transmembrane domains are indicated as TM1 through TM4, predicted with SOSUI (Hirokawa *et al.*, 1998). Identical residues are shown with a black background. **(A)** The sequences for yeast subunit 3 (Sc.3) and subunit 11 (Sc.11) are compared with Anc.3-11. A magenta background indicates residues that are identical only between subunit 3 and Anc.3-11, while a green background indicates identity between 11 and Anc.3-11. Glutamic acid residues critical for V-ATPase enzyme function are highlighted in yellow (subunit 3 E137, subunit 11 E145, and Anc.3-11 E139). A blue arrow indicates the position of subunit 3 I21 and Anc.3-11 M22. A red arrow indicates the position of subunit 11 F20 and Anc.3-11 V15. **(B)** The sequence for yeast subunit 16 (labeled Sc. 16) is compared to Anc.16. The critical glutamic acid residues (subunit 16 E108 and Anc.16 E98) are highlighted in yellow.



Supplement S4

Protein sequence alignment of contemporary *S. cerevisiae* proteolipids with reconstructed Anc.3-11 and intermediate ancestors Anc.11, and Anc.3. Potential transmembrane domains are indicated as TM1 through TM4. Identical residues are shown against a black background. **(A)** Anc.3-11 is compared to intermediate ancestor Anc.11 and contemporary yeast subunit 11 (labeled Sc.11). Residues shared exclusively between Anc.11 and Sc.11 are highlighted in green. The red arrow indicates the location of substitution V15F (see main text). **(B)** Anc.3-11 is compared to intermediate ancestor Anc.3 and contemporary yeast subunit 3 (Sc.3). Residues shared by only Anc. 3 and and Sc.c are highlighted in magenta. The blue arrow indicates the location of substitution M22I (see main text).

Supplement S5

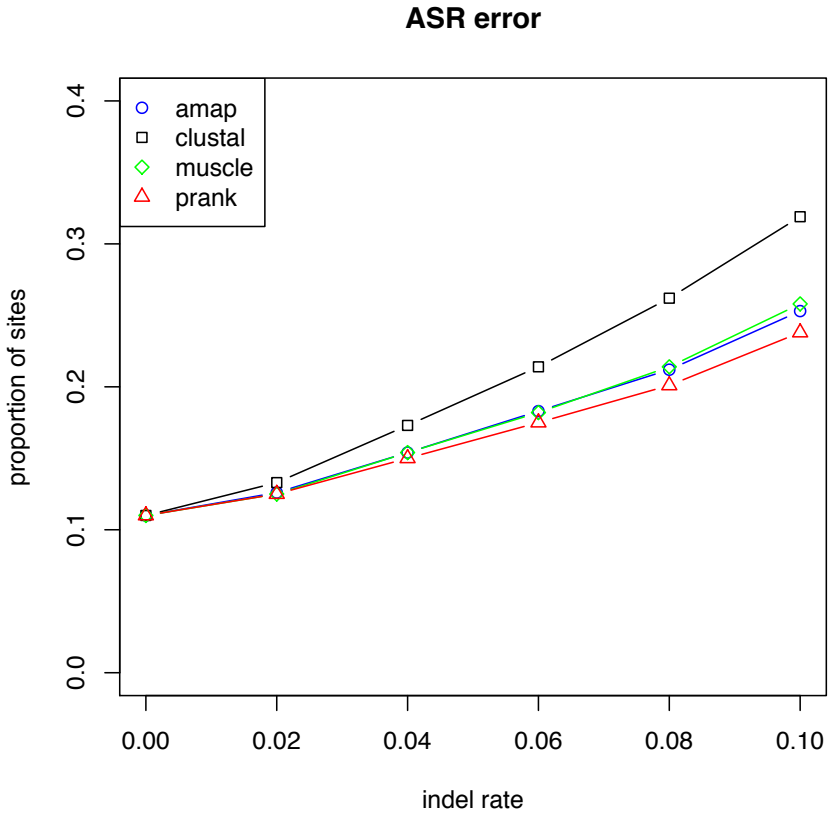
Yeast strains and vectors used in this study.

Supplemental Table 1. Yeast strains and plasmids used in this study

Strain	Genotype	Reference
SF838-1Dα	*MATα ura3-52 leu2-3,112 his4-519 ade6 pep4-3 gal2*	Rothman and Stevens (1986)
LGY113	SF838-1Dα; *vma3Δ::Kan^R*	Ryan *et al.* (2008)
LGY114	SF838-1Dα; *vma11Δ::Kan^R*	Ryan *et al.* (2008)
LGY115	SF838-1Dα; *vma16Δ::Kan^R*	Ryan *et al.* (2008)
LGY125	SF838-1Dα; *vma3Δ::Kan^R vma11Δ::Hyg^R*	This study
LGY143	SF838-1Dα; *vma3Δ::Kan^R vma11Δ::Hyg^R vma16Δ::Nat^R*	This study
LGY124	SF838-1Dα; *vma11Δ::Kan^R vma16Δ::Nat^R*	This study
LGY139	SF838-1Dα; *vma3Δ::Kan^R vma16Δ::Nat^R*	This study

Plasmid	Description	Reference
pRS415	*CEN, LEU2*	Simons *et al.* (1987)
pRS316	*CEN, URA3*	Sikorski and Hieter (1989)
pYEP351	*2μ, LEU2*	Hill *et al.* (1986)
pGF06	pRS316 *VPH1::GFP::HIS5*	Ryan *et al.* (2008)
pGF141	pRS415 *VPH1::GFP::HIS5*	This study
pGF140	pRS316 *prVMA3::Anc.3-11::3xHA::ADH::Nat^R*	This study
pGF139	pRS415 *prVMA16::Anc.16::3xHA::ADH::Nat^R*	This study
pGF252	pRS316 *prVMA3::Anc.3-11::3xHA::ADH* A41C, E139Q	This study
pGF253	pRS316 *prVMA16::Anc.16::3xHA::ADH* E98Q	This study
pGF254	pRS316 *VMA3*	This study
pGF499	pRS316 *VMA3* I21A	This study
pGF500	pRS316 *VMA11*	This study
pGF501	pRS316 *VMA11* F20A	This study
pGF241	pRS316 *prVMA3::Anc.3-11::3xHA::ADH* V15F	This study
pGF239	pRS316 *prVMA3::Anc.3-11::3xHA::ADH* A41S	This study
pGF240	pRS316 *prVMA3::Anc.3-11::3xHA::ADH* A120G	This study
pGF502	pRS415 *prVMA16::Anc.16(2-30Δ)::3xHA::ADH::Nat^R*	This study
pGF503	pRS316 *prVMA16::Anc.16::3xHA::ADH* I58L	This study
pGF504	pRS316 *prVMA16::Anc.16::3xHA::ADH* M77I	This study
pGF505	pRS316 *prVMA16::Anc.16::3xHA::ADH* K87R	This study
pGF213	pYEP351 *VMA21*	This study
pGF506	pRS316 *prVMA3::Anc.3::3xHA::ADH::Nat^R*	This study
pGF507	pRS415 *prVMA3::Anc.3::3xHA::ADH::Nat^R*	This study
pGF508	pRS316 *prVMA3::Anc.11::3xHA::ADH::Nat^R*	This study
pGF510	pRS316 *prVMA3::Anc.3-11::3xHA::ADH* V15A	This study
pGF512	pRS316 *prVMA3::Anc.3-11::3xHA::ADH* M22I	This study
pGF513	pRS316 *prVMA3::Anc.3-11::3xHA::ADH* S25T	This study
pGF514	pRS316 *prVMA3::Anc.3-11::3xHA::ADH* V38I	This study
pGF515	pRS316 *prVMA3::Anc.3-11::3xHA::ADH* A42G	This study
pGF517	pRS316 *prVMA3::Anc.3-11::3xHA::ADH* V45T	This study

pGF518	pRS316 *prVMA3::Anc.3-11::3xHA::ADH* M46L	This study
pGF519	pRS316 *prVMA3::Anc.3-11::3xHA::ADH* M46F	This study
pGF521	pRS316 *prVMA3::Anc.3-11::3xHA::ADH* I55L	This study
pGF523	pRS316 *prVMA3::Anc.3-11::3xHA::ADH* A61S	This study
pGF525	pRS316 *prVMA3::Anc.3-11::3xHA::ADH* K79L	This study
pGF528	pRS316 *prVMA3::Anc.3-11::3xHA::ADH* Y87F	This study
pGF529	pRS316 *prVMA3::Anc.3-11::3xHA::ADH* N88T	This study
pGF531	pRS316 *prVMA3::Anc.3-11::3xHA::ADH* H92Q	This study
pGF534	pRS316 *prVMA3::Anc.3-11::3xHA::ADH* F108Y	This study
pGF535	pRS316 *prVMA3::Anc.3-11::3xHA::ADH* T121Y	This study
pGF536	pRS316 *prVMA3::Anc.3-11::3xHA::ADH* A122M	This study
pGF537	pRS316 *prVMA3::Anc.3-11::3xHA::ADH* I132V	This study
pGF542	pRS316 *prVMA3::Anc. 3-11::3xHA::ADH* N159D	This study
pGF646	pRS415 *prVMA16::Sc.16(2-41Δ)::Anc.11(2-5Δ)::3xHA::ADH::Nat^R*	This study
pGF647	pRS415 *prVMA16::Anc.11(162-164Δ)::Sc.16(1-41Δ)::3xHA::ADH::Nat^R*	This study
pGF648	pRS415 *prVMA16::Sc.16(2-41Δ)::Anc.3-11(1-2Δ)::3xHA::Nat^R*	This study
pGF649	pRS415 *prVMA16::Anc.3-11(162Δ)::Sc.16(1-49Δ)::3xHA::Nat^R*	This study
pGF650	pRS415 *prVMA16::Sc.16(2-41Δ)::Anc.3(1-2Δ)::3xHA::Nat^R*	This study
pGF651	pRS415 *prVMA16::Anc.3(161-162Δ)::Sc.16(1-49Δ)::3xHA::Nat^R*	This study



Supplement S6

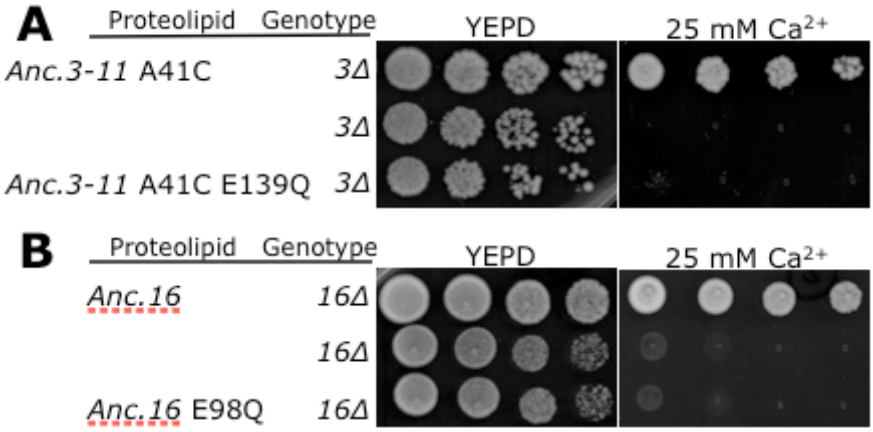
Ancestral sequence reconstruction (ASR) error as a function of insertion-deletion (indel) rate. Sequences were simulated on the phylogeny shown in S2, and aligned using four different algorithms: AMAP (Schwartz and Pachter 2007), Clustal (Thompson et al., 1994), MUSCLE (Edgar 2004), and Prank (Loytynoja and Goldman, 2008). We reconstructed the most-recent shared ancestors for the Anc.3-11 sequences and Anc.16 sequences. ASR was measured as the proportion of sites that incorrectly contained an indel character. Data is averaged over five replicates; standard error of the mean is very small (<0.001) and not rendered here.

A. S. Schwartz and L. Pachter. Multiple alignment by sequence annealing. *Bioinformatics*, **23**(2):e24–e29, 2007.

D. Thompson, D. G. Higgins, and T. J. Gibson. Clustal W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting position-specific gap penalties and weight matrix choice. *Nucleic Acids Research*, **22**(22):4673–4680, 1994.

Edgar. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research*, **32**(5):1792–1797, August 2004.

A. Loytynoja and N. Goldman. Phylogeny-aware gap placement prevents errors in sequence alignment and evolutionary analysis. *Science*, **320**(5884):1632–1635, June 2008.



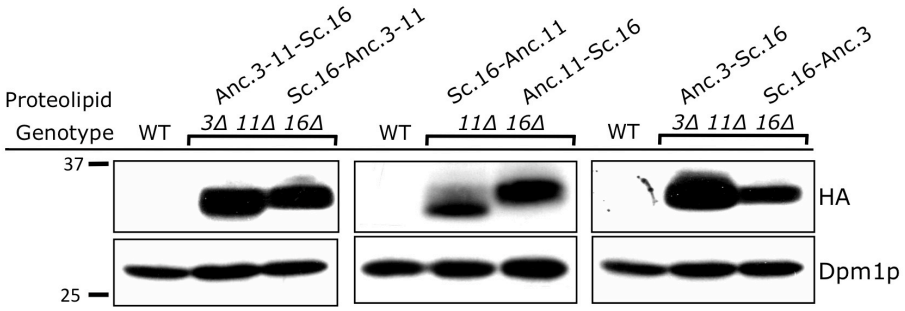
Supplement S7

Functional growth assays of Anc.3-11 and Anc.16 lacking critical glutamic acid residues in contemporary *S. cerevisiae*. Yeast were plated on permissive media (YEPD) and media buffered with 25 mM Ca^{2+}. **(A)** A particular allele of Anc.3-11 (A41C) was used to illustrate the dramatic growth difference when the E139Q substitution is introduced. Mutation of this essential residue causes a complete loss of V-ATPase function. **(B)** The corresponding glutamic acid residue within Anc.16 is also critical for V-ATPase enzyme function.

Proteolipid	Genotype		
			***vma3*Δ**
	***vma3*Δ**	***vma11*Δ**	***vma11*Δ**
Anc.3-11	+++	++	+++
Anc.3-11 V15F	++	++++	++
Anc.3-11 A41S	++	+	++
Anc.3-11 A120G	++	+	++
	***vma16*Δ**		
Anc.16	+++++		
Anc.16 I58L	+++++		
Anc.16 M77I	+++++		
Anc.16 K87R	+++++		
Anc.16 2-30Δ	+++++		

Supplement S8

Alternate ancestral states in Anc.3-11 and Anc.16 with posterior probability (pp) greater than 0.20 were tested in mutant yeast strains for V-ATPase function. We restricted our analysis to alternate states not located in sequence regions corresponding to lumenal portions of the protein (as predicted by SOSUI, Supplemental S3). The Anc.16 2-30Δ mutant contained 23 residues that had alternate states with greater than 0.20 pp, but nearly identical growth results. Alternate sites were introduced in Anc.3-11 and Anc.16 and their function was independently tested on media containing 25 mM Ca^{2+}. The number of "+" characters corresponds to the quality of growth. Wild-type yeast scored five pluses (+++++) and any yeast strain lacking a proteolipid subunit (*3*Δ, *11*Δ, *3*Δ*11*Δ, or *16*Δ) all scored zero under these conditions.



Supplement S9

Proteolipid gene fusions between yeast subunit 16 and Anc.3-11, Anc.3, or Anc.11 are stably expressed in modern yeast. Whole cell extracts were prepared from strains expressing gene fusion constructs, subjected to Western blot analysis, and probed using anti-HA antibodies or anti-Dpm1 antibodies (loading control). Extracts from wild-type yeast served as negative controls for the anti-HA antibody. The size of the nearest molecular marker (kilodaltons) is shown on the left.