Supporting Information

Purkanti and Thattai 10.1073/pnas.1407163112

SI Text

Sequence Clustering

We clustered sequences using default parameters in the graphical CLANS tool (1), which represents sequences as nodes and pairwise BLAST *P* values as attractive forces. We first examined the full alignment of 3,841 sequences. These split into seven robust clusters at the lenient *P*-value cutoff of 1*E*-1 (Fig. S1*D*). Three of these, classes A, B, and C, included sequences from multiple eukaryotic supergroups; the remaining were supergroup specific. LECA was the root species of class A and B dynamins; the last common ancestor of amoebozoans and archaeplastids was the root species of class C dynamins.

We next split the multiple-sequence alignment into eight distinct evolutionary segments, as described in the following section. Retaining only the 3,417 sequences from classes A, B, and C, we repeated the clustering analysis separately for each evolutionary segment (Fig. S1 *A*–*C*). Our goal was to find clusters of proteins that had high sequence identity across the corresponding segment, so the entire sequence stretch could be replaced by a single letter corresponding to the cluster label. At a lenient initial *P* value we found at most one cluster in each superclass containing proteins from multiple eukaryotic supergroups and therefore ancient. As the *P* value was made more stringent, derived clusters split away from this ancient cluster, which itself eventually disintegrated. This meant that, rather than using an arbitrary threshold, we could define clusters in an unsupervised manner over a range of *P* values. We labeled derived clusters by the *P* values at which they first emerged and the ancient cluster by the final *P* value just before it disintegrated. We retained clusters that persisted over a factor of 10 variation in *P* value and that contained sequences from more than one genus. All clusters were assigned unique numerical codes such as [segment.label] = [7.1]. Some were additionally assigned segment-wise letters, to aid brevity in the text and figures. The ancestral cluster at each segment was labeled by its functional superclass (e.g., 2A and 2B); clusters that spanned multiple eukaryotic groups or other relevant lineages were labeled with distinct letters (e.g., 1P, 2Q, and 3R), whereas most supergroup-specific clusters were labeled with a colon (:); gaps were labeled with a dash (-) (Table S1 and Dataset S1, sheets 1 and 2).

We assigned signatures as follows: Given a protein in the dataset, we split it along evolutionary segment boundaries. We examined the resulting subsequence at each segment and checked which cluster it was a member of. We then replaced the full amino acid sequence at each segment with the corresponding cluster label. This produced an eight-letter signature for each protein. We reconstructed ancestral variants by Dollo parsimony (2), as described in the main text: If two proteins had the same letter at the same segment, their last common ancestral protein was also assigned that letter. This representation in terms of short signatures is useful only if sequences within a cluster have high pairwise identity. Across segments, 223/238 clusters had a median pairwise identity >35% (Dataset S1, sheet 2). This means two proteins with the same letter at a segment typically have >35% sequence identity across that segment.

Evolutionary Segments

We want to split the alignment into a number of contiguous segments, so that parsimony can be usefully applied to the sequences at each segment. If the segments are too short, convergence is likely and parsimony is not valid. If the segments are too long, only recently diverged proteins will have a high sequence identity across the whole stretch, so very ancient proteins will fail to be reconstructed. Our approach is to start with small segments and grow them into longer segments. Imagine that we start off with a short segment of length *W*. If this were the final segment, we would feed in the corresponding sequences into CLANS, which uses BLAST *P* values based on BLOSUM62 distances (1, 3) to cluster sequences by similarity. Suppose we find that two contiguous segments, each of length *W*, contain sequences with identical pairwise BLOSUM62 distances. In that case, any pair of proteins that were similar across the first segment would also be similar across the second segment, so the two CLANS clustering analyses would be redundant. Such segments can be combined into a single longer segment with no loss of information. We continue this procedure until we encounter a segment that has a very different set of pairwise BLOSUM62 distances. In that case, we end the growth of one long segment and start the growth of the next. We start at the N terminus and repeat until we reach the C terminus.

We implement this as follows. From the full 556-residue multiple-sequence alignment, we extract the subalignment corresponding to length-*W* sequences starting at position *i*. We then calculate the list of all pairs of BLOSUM62 distances between these sequences, excluding those completely represented by gaps. Given two such lists starting at positions *i* and *j*, we can quantify their similarity, using Pearson's correlation coefficient over corresponding pairs. Two positions with high correlation coefficients have similar pairwise BLOSUM62 distances and would therefore produce similar CLANS clusters. We want correlations only for nonoverlapping windows, because windows with shared residues will be trivially correlated; we therefore split the protein into successive windows of length *W* and calculate correlations only for pairs $\{i, j\}$, where *i* and *j* run over multiples of *W*. This produces a symmetric correlation matrix P_{ij}, in which values close to 1 indicate that subsequences starting at positions *i* and *j* will have similar CLANS clusters (lower triangular portion, Fig. 1*B*).

We next want to find out where the breaks between evolutionary segments should be placed. Rather than using the correlation matrix P_{ij} directly, we opt to use a smoothed distance matrix based on moving windows, which has better dynamic range and spatial resolution. Regarding the *i*th column of P_{ij} as representing the coordinates of the *i*th residue in some vector space, we can calculate the squared Pythagorean distance between columns *i* and *j*. This produces a distance matrix Q_{ij}, in which values close to 0 indicate strong similarity. We can repeat this procedure for shifted windows, with *i* and *j* running over values $w = 0, \ldots, W-1$ modulo *W*, so that each residue from *W* to $557-W$ is covered by *W* consecutive windows. By averaging the corresponding matrix values, we obtain smoothed versions of the correlation and distance matrices P'_{ij} and Q'_{ij}, where *i* and *j* now run from *W* to $557-W$.

In this representation, evolutionary segments correspond to contiguous sets of residues with small pairwise distances. For every potential segment running over residues *i*–*j*, we can extract the corresponding square submatrix of the smoothed distance matrix Q'_{ij} and calculate the fraction $R_{ij}(q)$ of its entries that lie below some threshold value *q*. By manually adjusting the values of *q* and R_{ij}, we obtain successive nonoverlapping segments. Our analysis was carried out using a window size $W = 10$ and segments within which $R_{ij} = 0.75$ of pairwise distances were below a threshold of $q = 1.2$ (upper triangular portion, Fig. 1*B*). These parameters were chosen to strike a balance between having sufficiently long segment lengths to rule out convergence and

having sufficiently many segments to sample nonuniform clustering properties across the protein sequence. This produced eight well-defined segments ranging from 22 to 118 residues in length, much smaller than domains.

To understand why segments emerge at this length scale, we compared the positions of evolutionary segments with secondary-structural elements in crystal structures of rat (PDB ID: 3ZVR) and human (PDB ID: 3SNH) dynamin 1. We first split the protein into α-helices, β-sheets, and intervening gaps as seen in the crystal structures and mapped the segments from the alignment onto the actual protein sequences. For the left and right boundaries of each segment, we determined the separation in residues to the nearest boundary of a secondary-structural element and recorded the minimum of these two values as the "true separation." We repeated this procedure, using 1,000 shuffled proteins in which secondary-structural elements and gaps were randomly permuted, each time recording a "shuffled separation." We then calculated the *P* value: the fraction of shuffled separations less than or equal to the true separation. If evolutionary segments were uncorrelated with secondary structure, the true separation would be similar to those seen in shuffled proteins, resulting in a *P* value close to 1. In fact, the observed *P* values are much lower: rat dynamin 1 (4), $P = 0.008$; human dynamin 1 (5), $P = 0.106$. This suggests that evolutionary segments correspond to the scale of protein secondary structure.

Benchmarking Segment-Based Parsimony

We want to compare the predictions of our method against those of a maximum-likelihood analysis. This requires a protein dataset whose phylogenetic structure is known, so we can benchmark performance in terms of the likelihood of reconstructing the correct phylogeny. Here we focus on the class A dynamins of land plants, whose mutual relationships were reported by Miyagishima et al. (6). Land plants have three varieties of class A dynamins (parentheses give gene names from ref. 6 as well as Uniprot IDs): a mitochondrial division variant (MID), with signature [AAAAAA::] (e.g., AtDRP3A/Q8S944); a PH-domain–containing vesicle scission variant (VES), with signature [:W:-:XYZ] (e.g., AtDRP2A/Q9SE83); and a phragmoplastin (PHR) responsible for cell plate formation, with signature [AVAA:XYZ] (e.g., AtDRP1A/P42697). As an outgroup for maximum-likelihood analyses, we use the plant Mx-like dynamin (MX) with signature [BB:::::] (e.g., AtDRP4C/Q9ZP55). Our dataset has 17 species of land plants containing one or more copies of each of these four variants, giving 196 proteins in total. Our segment-based parsimony method relies on knowing the signatures of these dynamins, as well as those of others spread across the backbone tree of eukaryotic species (Fig. 3). For the purposes of the present analysis, we observe that metazoans have a mitochondrial dynamin with signature [AAAAAAAA]. Based on these signatures alone, we make several nontrivial predictions. We test each of these predictions, using appropriate RAxML maximum-likelihood analyses (*Methods: Maximum-Likelihood Analysis*):

i) We predict a unique and robust phylogeny in which plant phragmoplastins and vesicle dynamins are the closest relatives (Fig. S3*A*, *Top*), as reported previously (6). Our RAxML analysis on the full alignment of 196 proteins generates precisely the same tree with 100/100 bootstrap support for all nodes (Fig. S3*B*, *Top*).

ii) Enforcing the overall phylogeny at each segment independently, we predict that sequences labeled with the letters 2V, 2W, 6X, 7Y, and 8Z should form separate monophyletic clades (Fig. S3*A*, *Bottom*). For example, this is equivalent to claiming that the last common protein ancestor of the cluster of sequences labeled by 2V has only 2V-type descendants. In contrast, we predict that the clusters labeled A at each segment are paraphyletic. This means that the last common protein ancestor of sequences labeled by A also has descendants with other labels (Fig. S3*A*, *Bottom*). We test this as described in Fig. 2 and the associated text; results are summarized in Dataset S1, sheet 2. The RAxML analysis assigns predicted monophyletic clusters the following high monophyly bootstrap support: 2V, 856/1,000; 2W, 996/1,000; 6X, 900/1,000; 7Y, 374/1,000; and 8Z, 934/1,000. In contrast, the clusters labeled by A at each of the eight segments are robustly paraphyletic, with little or no monophyly bootstrap support: 1A, 104/1,000; 2A, 0/1,000; 3A, 0/1,000; 4A, 13/1,000; 5A, 0/1,000; 6A, 0/1,000; 7A, 70/1,000; and 8A, 0/1,000.

iii) Our parsimony-based reconstruction (Fig. S3*A*, *Top*) immediately suggests that the C-terminal portion of the protein is phylogenetically informative but the N-terminal portion is not; that is, we predict that segments 6–8 are necessary to infer the correct phylogeny. To test this, we ran RAxML analyses on a large number of subsequences of the protein, with 100 bootstrap replicates for each run. In each case, we asked how many of these replicates generated the correct phylogeny, grouping phragmoplastins with vesicle dynamins, rather than the incorrect phylogeny grouping phragmoplastins with mitochondrial dynamins (Fig. S3*B*, *Bottom*; the replicates that do not fall into these two categories are poorly resolved, with low bootstrap support for all nodes). We first examined two truncated series, one starting at the N terminus and getting longer and another starting at the C terminus and getting longer (Fig. S3*C*). As predicted, the C-terminal series recovers the correct phylogeny with 99/100 bootstrap support on a sequence of just 100 residues, whereas the N-terminal series achieves similar performance only on sequences of 400 residues or more. From these data it appears that segment 7 is the most phylogenetically informative, whereas segment 2 is the least. To see why this is the case, we again ran a series of RAxML analyses, this time centered within each of segments 2 and 7, but with increasing length up to the full segment length (Fig. S3*D*). We compared these results to those on random subsamples of the full protein sequence for the same lengths. Strikingly, segment 2 recovers precisely the wrong phylogeny with increasing length, due to the high N-terminal divergence of the vesicle dynamins: This is a case of long-branch attraction, which forces cluster 2W to the center of the tree in Fig. 2*B*. In contrast, both segment 7 and the random subsamples recover the correct phylogeny with increasing length. However, whereas the random subsamples achieve a performance of 87/100 at a length of 100 residues, segment 7 achieves a performance of 88/100 at a length of just 20 residues. It is only once we reduce the sequence length below 15 residues that the performance truly suffers due to accidental convergence, and no phylogeny is recovered with any significant support.

iv) An alternative, less parsimonious, hypothesis for the patterns seen in Fig. S3*A* is that the vesicle dynamin is indeed a highly diverged variant and that the phragmoplastin is the result of fusion between the N-terminal region of a mitochondrial dynamin and the C-terminal region of a vesicle dynamin. If fusion were the case, we would have predicted that subsequences recovered distinct phylogenies with high bootstrap support (as seen for segments 2 and 7), but also that the full sequence containing conflicting phylogenetic components would have lower support than these subsequences for any given phylogeny. Instead, we see that, with increasing protein length, the phylogeny of Fig. S3*B* (*Top*) is recovered with increasing support. The monotonic nature of the bootstrap support curves from the N-terminal and C-terminal series is completely inconsistent with the fusion hypothesis.

Taken together, these results show that multiple nontrivial predictions made using segment-based parsimony on this benchmarking

dataset are fully validated using a maximum-likelihood analysis. Parsimony also easily scales to very large datasets for which maximum likelihood would be computationally intractable. The most important feature of our approach is that phylogenetic signal is spatially represented across the protein sequence. This allows us to reconstruct ancestral sequences to a desired confidence level [e.g., >35% identity, comparable to that between proteins of similar function (7)] and to explicitly indicate which ancestral segments cannot be reconstructed. Therefore, if we are able to find present-day proteins with the same signature as some reconstructed ancestral variant, we can infer the function of the ancestral variant with high confidence. The remarkable fact, and one of our central results, is that such "living fossils" do indeed exist.

Routes of Gene Duplication

We consider the evolution of a small haploid population in which neutral mutations arise rarely but are fixed quickly, so the population is genetically monomorphic between successive fixation events (i.e., the $N\mu << 1$ limit, where N is the population size and μ is the mutation rate). We model a gene as having two domains assigned to potential functions X and Y, respectively; nonfunctional alleles are labeled O. We examine all routes, taking a single initial gene XO to two final specialized genes $XO + OY$. We assume that purifying selection maintains function X initially and maintains both functions X and Y once the latter is discovered (Fig. 4*E*). Gain-of-function mutations taking O to X or Y are discovered at rate α and instantaneously fixed. We expect α to be independent of population size because it is dominated by a neutral search through sequence space; $1/\alpha$ therefore represents a reference timescale. Loss-of-function mutations taking X or Y to O are fixed at rate β. We take the large-β limit so that genes rapidly reach the most degraded states compatible with the necessary functions; for example, $XY + XO$ could degrade to $XY + OO$ or $OY + XO$. Gene duplications and deletions are fixed at respective rates γ and δ. We assume that redundant gene copies have a metabolic fitness cost, so γ decreases and δ increases with increasing population size, and the chance of having more than two gene copies is negligible. Our analysis will break down at a population size where duplicate genes can accumulate mutations before fixation (8). These dynamics are expressed in the following transition diagram, where the return path to $XY + OO$ is possible from any instance of $XO + OY$, but shown only once for clarity:

$$\begin{array}{ccccc} XO+OY & \overset{\gamma/2}{\leftarrow} & XY & \underset{\delta}{\overset{\gamma/2}{\rightarrow}} & XY+OO \underset{\alpha}{\overset{\alpha}{\rightarrow}} XO+OY \\ & & \uparrow\alpha & & \alpha\uparrow \\ & & XO & \underset{\delta}{\overset{\gamma}{\rightarrow}} & XO+OO \overset{\alpha}{\rightarrow} XO+OY \end{array}$$

The expected time spent in futile cycles is given by

$$\tau_{XO} = \frac{\gamma\delta}{\alpha(2\alpha+\delta+2\gamma)}\left(\frac{1}{\alpha+\gamma}+\frac{1}{2\alpha+\delta}\right),$$

$$\tau_{XY} = \frac{\delta}{2\alpha+\delta}\left(\frac{1}{\gamma}+\frac{1}{\alpha+\delta}\right).$$

The probabilities and conditional mean first-passage times of possible routes are then as follows (futile cycles of gene duplication and deletion are shown as overbars):

Route	Probability	MFPT
$\overleftrightarrow{XO} \rightarrow XO + OO \rightarrow XO + OY$	$\frac{\gamma}{2\alpha+\delta+2\gamma}$	$\tau_{XO} + \frac{1}{\alpha+\gamma} + \frac{1}{2\alpha+\delta}$
$\overleftrightarrow{XO} \rightarrow XY$	$\frac{2\alpha+\delta}{2\alpha+\delta+2\gamma}$	$\tau_{XO} + \frac{1}{\alpha+\gamma}$
$\overleftrightarrow{XO} \rightarrow XO + OO \rightarrow XY + OO$	$\frac{\gamma}{2\alpha+\delta+2\gamma}$	$\tau_{XO} + \frac{1}{\alpha+\gamma} + \frac{1}{2\alpha+\delta}$
$\overleftrightarrow{XY} \rightarrow XO + OY$	$\frac{\alpha+\delta}{2\alpha+\delta}$	$\tau_{XY} + \frac{1}{\gamma}$
$\overleftrightarrow{XY} \rightarrow XY + OO \rightarrow XO + OY$	$\frac{\alpha}{2\alpha+\delta}$	$\tau_{XY} + \frac{1}{\gamma} + \frac{1}{\alpha+\delta}$
$\overleftrightarrow{XY + OO} \rightarrow XY \rightarrow XO + OY$	$\frac{\delta}{2\alpha+\delta}$	$\tau_{XY} + \frac{1}{\gamma} + \frac{1}{\alpha+\delta}$
$\overleftrightarrow{XY + OO} \rightarrow XO + OY$	$\frac{2\alpha}{2\alpha+\delta}$	$\tau_{XY} + \frac{1}{\alpha+\delta}$.

We refer to states XO and $XO + OO$ as monofunctional (M), states XY and $XY + OO$ as bifunctional (B), and the state $XO + OY$ as specialized (S). The expected time spent in a monofunctional state when starting from XO and the probabilities to exit into either a specialized or a bifunctional state are

$$\frac{1}{2\alpha} < \tau^M_{M\rightarrow S} < \frac{3}{2\alpha} \quad P_{M\rightarrow S} = \frac{\gamma}{2\alpha+\delta+2\gamma}$$
$$\frac{1}{2\alpha} < \tau^M_{M\rightarrow B} < \frac{1}{\alpha} \quad P_{M\rightarrow B} = \frac{2\alpha+\delta+\gamma}{2\alpha+\delta+2\gamma}.$$

The expected times spent in the bifunctional state when entering from XO or returning from $XO + OY$ or in the specialized state before returning to the bifunctional state are

$$\frac{1}{2}\left(\frac{1}{\gamma}+\frac{1}{\alpha+\delta}\right) < \tau^B_{M\rightarrow B\rightarrow S} < 2\left(\frac{1}{\gamma}+\frac{1}{\alpha+\delta}\right)$$
$$\frac{1}{\alpha+\delta} < \tau^B_{S\rightarrow B\rightarrow S} < 2\left(\frac{1}{\gamma}+\frac{1}{\alpha+\delta}\right) \quad \tau^S_{S\rightarrow B} = \frac{1}{\alpha}.$$

Two types of trajectories are possible (arrows are labeled with approximate lifetimes):

$$M \overset{1/\alpha}{\rightarrow} S \overset{1/\alpha}{\rightarrow} B \overset{1/\gamma+1/(\alpha+\delta)}{\rightarrow} S \overset{1/\alpha}{\rightarrow} B \rightarrow \ldots$$
$$M \overset{1/\alpha}{\rightarrow} B \overset{1/\gamma+1/(\alpha+\delta)}{\rightarrow} S \overset{1/\alpha}{\rightarrow} B \overset{1/\gamma+1/(\alpha+\delta)}{\rightarrow} S \rightarrow \ldots.$$

We now take the reasonable limit $\alpha << \delta$ where gain-of-function mutations are rarer than deletions of redundant genes; this strongly favors the second of the two trajectories shown above. If we approximate each transition as a single effective step whose duration is exponentially distributed, we get the following transition probabilities:

$$P_{M\rightarrow M}(t) = e^{-\alpha t}$$

$$P_{M\rightarrow S}(t) = (1 - e^{-\alpha t}) - \frac{\alpha}{\alpha+\omega}\left(1 - e^{-(\alpha+\omega)t}\right)$$

$$P_{M\rightarrow B}(t) = \frac{\alpha}{\alpha+\omega}\left(1 - e^{-(\alpha+\omega)t}\right)$$

$$P_{S\to S}(t) = 1 - \frac{\alpha}{\alpha+\omega}\left(1 - e^{-(\alpha+\omega)t}\right)$$

$$P_{S\to B}(t) = \frac{\alpha}{\alpha+\omega}\left(1 - e^{-(\alpha+\omega)t}\right)$$

$$P_{B\to S}(t) = \frac{\omega}{\alpha+\omega}\left(1 - e^{-(\alpha+\omega)t}\right)$$

$$P_{B\to B}(t) = 1 - \frac{\omega}{\alpha+\omega}\left(1 - e^{-(\alpha+\omega)t}\right).$$

There are only two parameters: $1/\alpha$ is the gain-of-function timescale; and $1/\omega = 1/\gamma + 1/\delta$ is the subfunctionalization timescale, equivalent to a cycle of gene duplication and deletion. Note that the dependence of $1/\omega$ on the population size is not obvious. Starting from M, for short times $t << 1/(\alpha + \omega)$, given that a transition has occurred it is most likely to be into the bifunctional state. For long times $t >> 1/\alpha$, the system reaches an equilibrium between the specialized and bifunctional states, depending only on the ratio $\omega/(\alpha + \omega)$.

We now follow this process on a phylogenetic tree (Fig. S5). Only when ω is very large do we find that the ancestor is likely to be monofunctional (for an ancient ancestor) or specialized (for a recent ancestor). In all other situations the ancestor is most likely to be bifunctional. For large α,ω the ancestor is sampled from the equilibrium distribution and therefore is more likely to be bifunctional for $\alpha > \omega$. For fixed ω but $\alpha < 1$ the ancestor is more likely to be bifunctional as we decrease α: Because it is difficult to escape from a specialized state, the only option is to transition early into a bifunctional state. For very small α, ω this analysis reduces to parsimony, and the ancestor is overwhelmingly likely to be bifunctional no matter how ancient it is.

1. Frickey T, Lupas A (2004) CLANS: A Java application for visualizing protein families based on pairwise similarity. *Bioinformatics* 20(18):3702–3704.
2. Farris JS (1977) Phylogenetic analysis under Dollo's law. *Syst Zool* 26(1):77–88.
3. Altschul SF, et al. (1997) Gapped BLAST and PSI-BLAST: A new generation of protein database search programs. *Nucleic Acids Res* 25(17):3389–3402.
4. Ford MGJ, Jenni S, Nunnari J (2011) The crystal structure of dynamin. *Nature* 477(7366): 561–566.
5. Faelber K, et al. (2011) Crystal structure of nucleotide-free dynamin. *Nature* 477(7366): 556–560.
6. Miyagishima SY, Kuwayama H, Urushihara H, Nakanishi H (2008) Evolutionary linkage between eukaryotic cytokinesis and chloroplast division by dynamin proteins. *Proc Natl Acad Sci USA* 105(39):15202–15207.
7. Orengo CA, et al. (1997) CATH—a hierarchic classification of protein domain structures. *Structure* 5(8):1093–1108.
8. Lynch M, Force A (2000) The probability of duplicate gene preservation by subfunctionalization. *Genetics* 154(1):459–473.

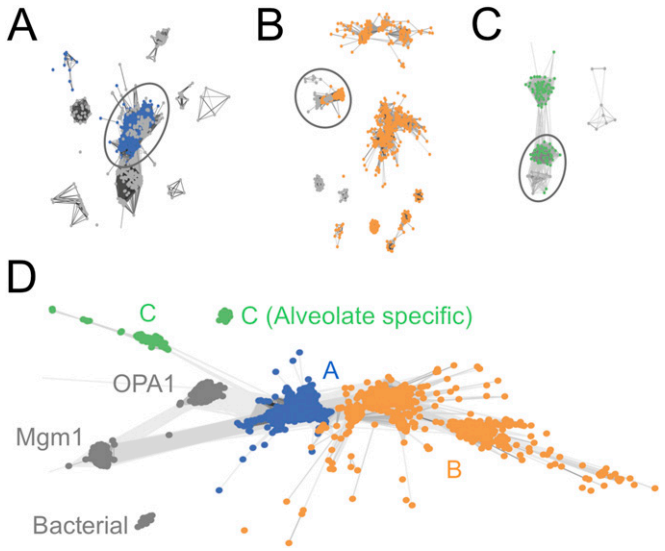

Fig. S1. Clustering by sequence similarity. CLANS clustering: Each node is a single sequence, and edges represent sequence similarity. (*A*–*C*) Shown are snapshots of the clustering procedure for segment 2 at an intermediate *P* value. Ancient clusters (oval) are those containing proteins from multiple eukaryotic supergroups. (*A*) Class A, *P* value = 6*E*-28 (root species, LECA; blue, opisthokont side of root; gray, others). (*B*) Class B, *P* value = 1*E*-16 (root species, LECA; orange, opisthokont side of root; gray, others). (*C*) Class C, *P* value = 9*E*-14 (root species, last common ancestor of amoebozoans and archaeplastids; green, archaeplastid side of root; gray: others). (*D*) Clustering of the full alignment of 3841 sequences across 556 residues. These split into seven robust clusters at the lenient *P* value cutoff of 1*E*-1; for clarity in this 2D projection we have shown a snapshot at *P* value 1*E*-27. Three of these clusters, which we term functional superclasses, include sequences from multiple eukaryotic supergroups: Class A (blue, 1,891 sequences) contains mitochondrial and peroxisomal division dynamins, vesicle dynamins, and phragmoplastins; class B (orange, 1,390 sequences) contains myxovirus-resistance-like dynamins; class C (green, 136 sequences) contains cytokinetic and chloroplast dynamins; an alveolate-specific set (green, 16 sequences) groups with class C in a phylogenetic analysis (Fig. S2) and is labeled "C (Alveolate specific)" in Dataset S1. The remaining clusters (gray) include the mitochondrial fusion dynamins Mgm1 (fungal; 252 sequences) and OPA1 (metazoan; 142 sequences) and a bacterial set (14 sequences).

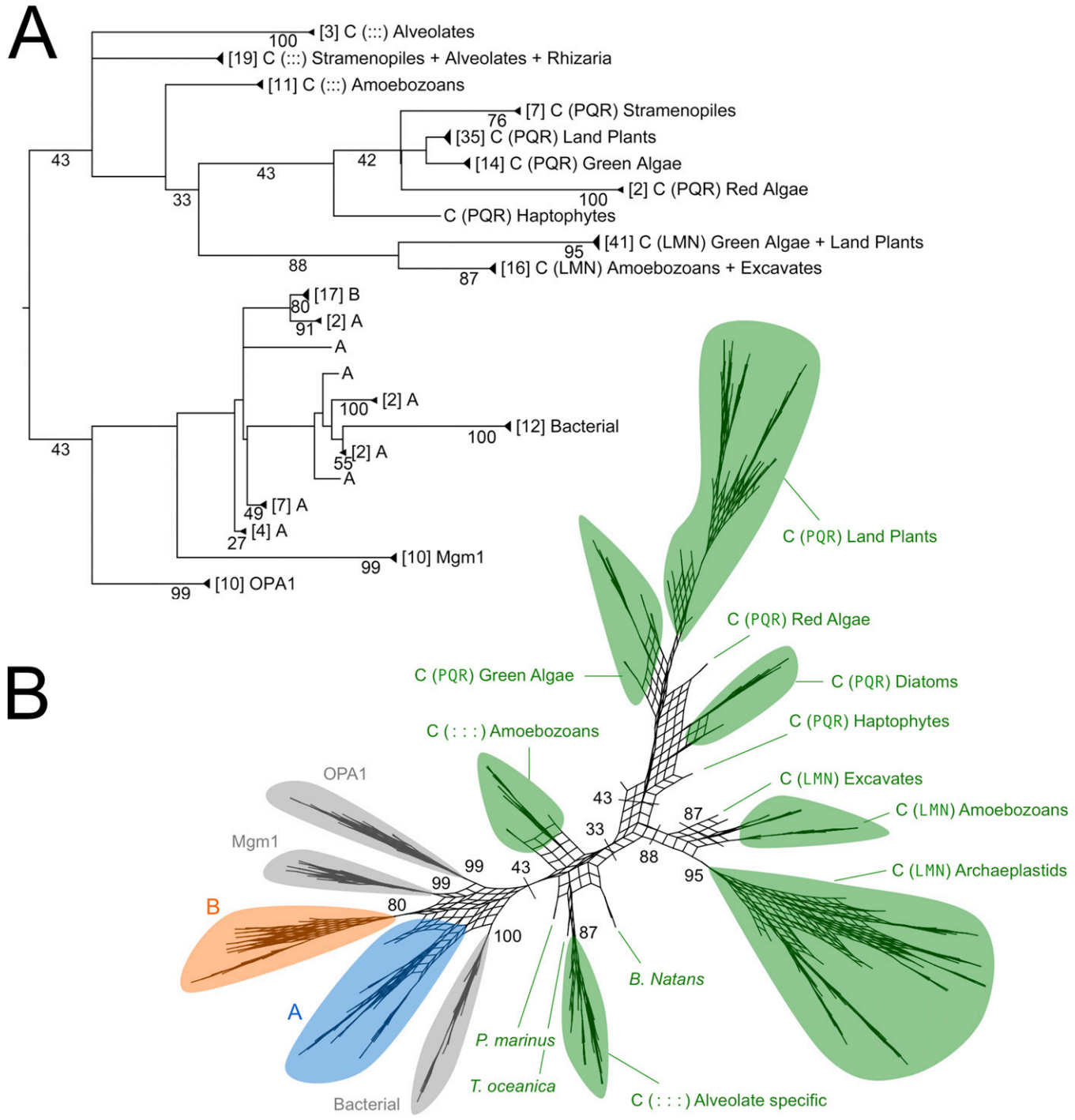

Fig. S2. Phylogeny of GTPase domains of major dynamin variants. Shown is the result of a RAxML phylogenetic analysis of concatenated segments 1–3, covering the dynamin GTPase domain. This analysis includes the entire set of class C (including C: Alveolate specific) and "bacterial" dynamins, along with 10 randomly sampled ancestral class A and Class B dynamins on each side of LECA and 10 randomly sampled dynamins from each of the "Mgm1" and "OPA1" groups. (*A*) Rooted consensus tree showing bootstrap values above 30%. (*B*) The consensus split network (1) summarizing 100 bootstrap replicate trees. Clusters with strong monophyletic bootstrap support appear with thin stems in the consensus split network. The cytokinetic class C dynamins (LMN; bootstrap support 88%) (2) are strongly monophyletic and form a sister group to the plastid division class C dynamins (PQR; bootstrap support 43%). Both of these appear descended from an ancestor whose sequence has been lost (labeled in the main text as ???). The class A cluster is paraphyletic in the traditional phylogram representation in *A*: Both the class B and bacterial groups branch from within it. In the consensus split network this paraphyletic cluster appears to emerge from a "gall." This reflects the uncertainty in finding the closest class A relatives of the subbranching groups.

1. Huson DH, Bryant D (2006) Application of phylogenetic networks in evolutionary studies. *Mol Biol Evol* 23(2):254–267.
2. Shiflett AM, Johnson PJ (2010) Mitochondrion-related organelles in eukaryotic protists. *Annu Rev Microbiol* 64:409–429.

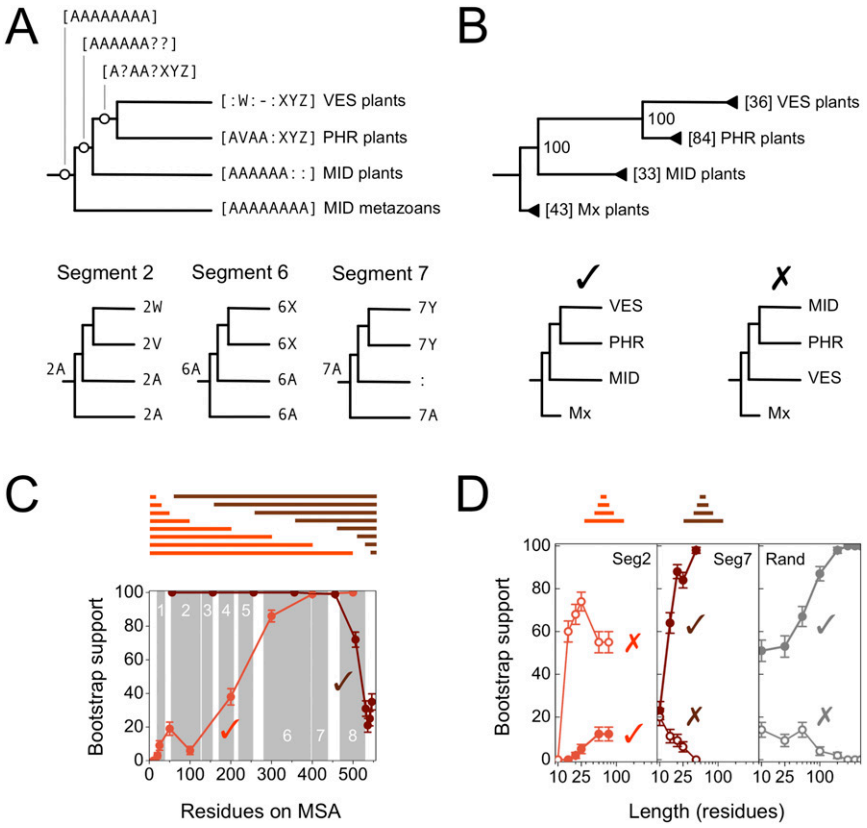

Fig. S3. Benchmarking evolutionary segments. (*A*) (*Top*) Phylogeny of class A dynamins of land plants, based on parsimony. VES, vesicle scission; PHR, phragmoplastin; MID, mitochondrial division. By comparing the mitochondrial division dynamins of metazoans and land plants we establish the ancestral states at each node. (*Bottom*) The same phylogeny must apply for each segment separately. (*B*) (*Top*) The RAxML tree for the full benchmarking dataset, with plant Mx proteins as the outgroup. All branches and leaves are recovered with 100/100 bootstrap support; numbers in brackets indicate the number of sequences in each collapsed clade. (*Bottom*) Correct and wrong phylogenies are indicated with a tick mark and a cross, respectively. (*C*) Results of RAxML analyses on a truncated series of protein subsequences. Error bars show binomial confidence intervals. The graph covers the full 556-residue multiple sequence alignment (MSA); gray areas indicate the eight evolutionary segments. Each curve measures the fraction of bootstrap replicates on a truncated sequence that recover the correct phylogeny. Light brown: Sequences starting at the N terminus and getting longer (plotted from the left; longer segments at higher *x* values). Dark brown: Sequences starting at the C terminus and getting longer (plotted from the right; shorter segments at higher *x* values). E.g., the light brown datapoint at the *x* value of 100 measures the result for a 100-residue N-terminal variant running from positions 1 to 100 of the alignment; the dark brown datapoint at the *x* value of 457 measures the result for a 100-residue C-terminal variant running from positions 457 to 556 of the alignment. (*D*) Results of RAxML analyses on truncated subsequences centered at segment 2 (*Left*) (Seg2: light brown) or segment 7 (*Center*) (Seg7: dark brown), for a series of lengths up to the full segment length (*x* axis, logarithmic). (*Right*) (Rand; gray) The results for sequences of various lengths constructed by randomly subsampling residues from the full protein. Shown is how often the correct phylogeny (solid circles, tick marks) or the wrong phylogeny (open circles, crosses) is recovered.

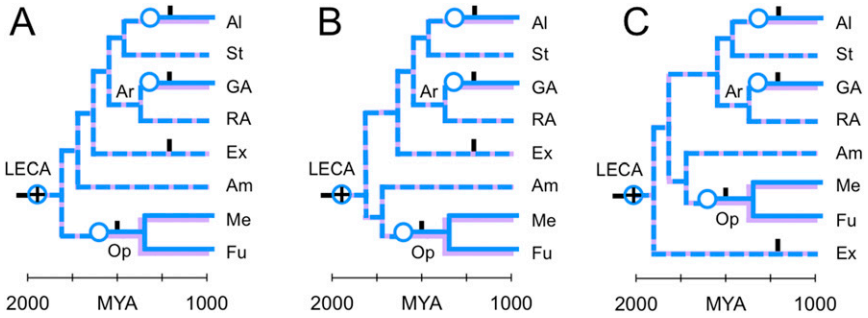

Fig. S4. Gene duplication scenarios with alternative rootings of the eukaryotic tree. Species labels: Al, alveolates; St, stramenopiles; Ar, archaeplastids; GA, green algae (including land plants); RA, red algae; Ex, excavates; Am, amoebozoans; Op, opisthokonts; Me, metazoans; Fu, fungi. Colors and symbols correspond to those used in Fig. 4. The opisthokonts, including fungi and metazoans, descend from one side of LECA. The archaeplastids and the SAR supergroup descend from the other side of LECA. The archaeplastids are the chloroplast-containing lineages, including red algae, green algae, and plants. The SAR supergroup split from the archaeplastids before chloroplast endosymbiosis: They include stramenopiles such as oomycetes and diatoms, alveolates such as ciliates, and rhizaria. Two other supergroups—amoebozoans and excavates—appear monophyletic but their placement is not resolved. In the main text, we use a tree in which amoebozoans and excavates are placed on the opposite side of LECA to the opisthokonts (1). Our results would not change if amoebozoans were grouped with opisthokonts (2), except that the term "last common ancestor of amoebozoans and archaeplastids" would then refer to LECA itself. Here we show the most parsimonious scenarios that explain the present-day distribution of bifunctional and specialized dynamins, for different rootings of the eukaryotic tree. (*A*) The scenario shown in Fig. 4*B*: Amoebozoans and excavates are placed on the opposite side of LECA to the opisthokonts. (*B*) Amoebozoans are grouped with the opisthokonts, and excavates are placed on the opposite side of LECA. (*C*) Excavates are placed as an outgroup, and amoebozoans are grouped with the opisthokonts. All these rootings predict a bifunctional ancestral dynamin that duplicated and specialized at least three independent times, in the ancestors of opisthokonts, green algae, and alveolates.

1. Parfrey LW, Lahr DJ, Knoll AH, Katz LA (2011) Estimating the timing of early eukaryotic diversification with multigene molecular clocks. *Proc Natl Acad Sci USA* 108(33):13624–13629.
2. Adl SM, et al. (2012) The revised classification of eukaryotes. *J Eukaryot Microbiol* 59(5):429–493.

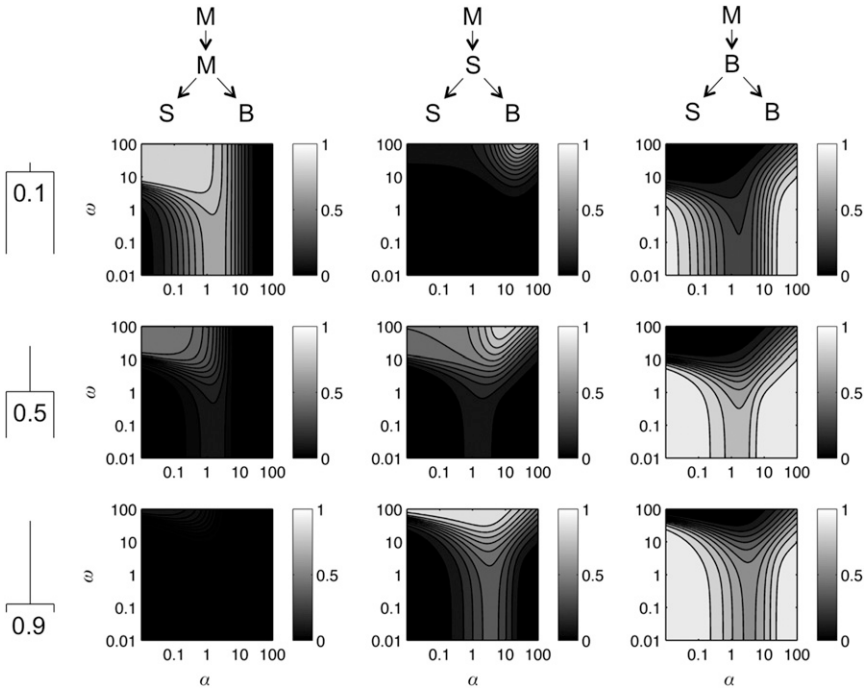

Fig. S5. Modeling gene duplication on a phylogenetic tree. Suppose the gene M (monofunctional) arose at time $t = 0$, and we find two species at time $t = 1$ with genotypes S (specialized) and B (bifunctional), respectively. The last common ancestor of these species must date to a time $0 < \text{TLCA} < 1$ (TLCA, time to last common ancestor) and must have had one of the genotypes M, S, or B. Assuming all three ancestral genotypes are equally likely a priori, we can calculate their Bayesian posterior probabilities for fixed α, ω, and TLCA. These values are shown for TLCA = 0.1, 0.5, and 0.9 and $0.01 \leq \alpha,\omega \leq 100$; lighter shading indicates higher probabilities.

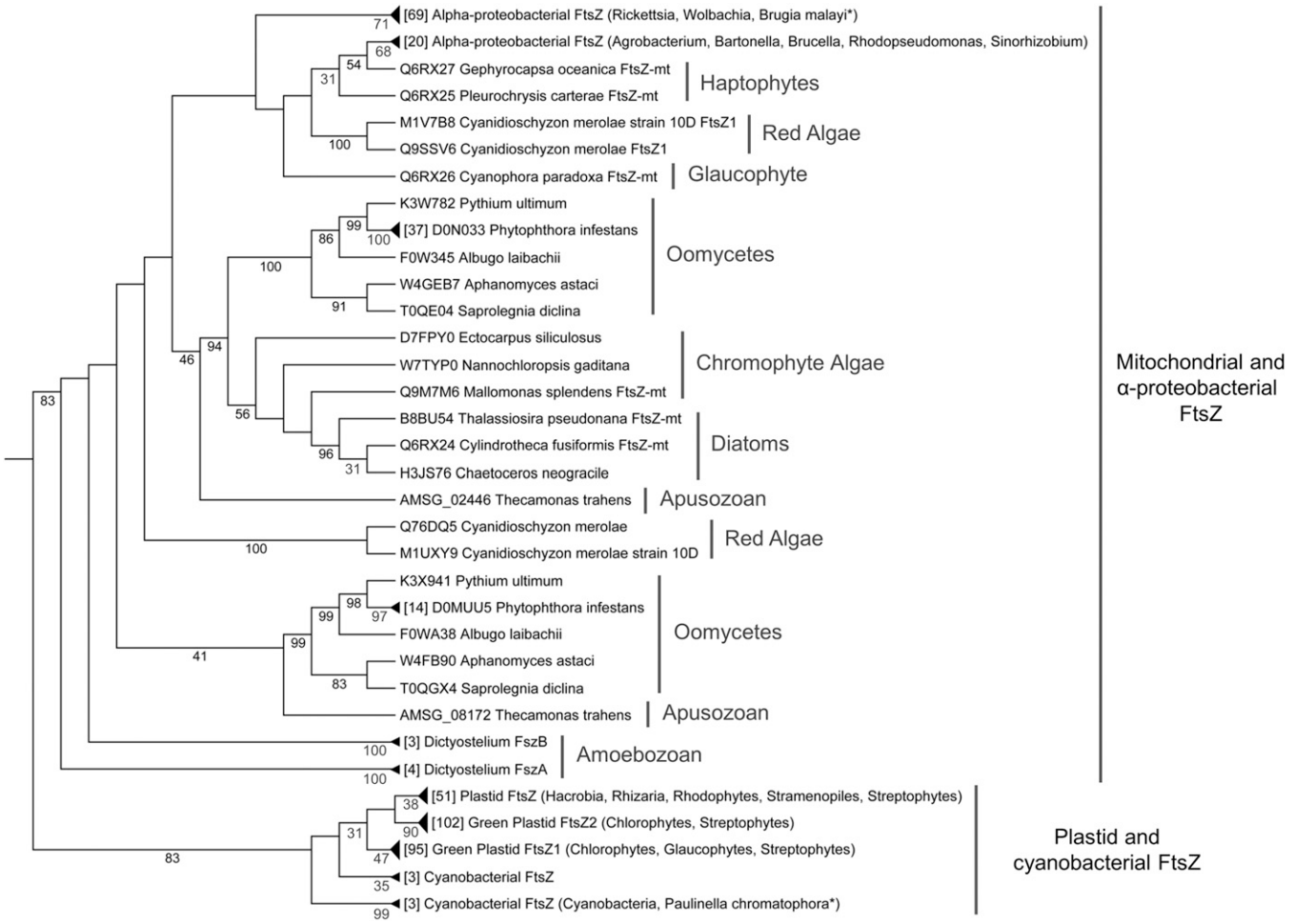

Fig. S6. Extant eukaryotic lineages that retain mitochondrial FtsZ. To detect mitochondrial FtsZ homologs, we repeated the analysis reported in Kiefel et al. (1). We started with a set of seven mitochondrial FtsZ protein sequences from the amoebozoan *Dictyostelium discoideum* (Q9GPZ7), the glaucophyte *Cyanophora paradoxa* (Q6RX26), the red alga *Cyanidioschyzon merolae* (Q9SSV6), the haptophytes *Gephyrocapsa oceanica* (Q6RX27) and *Pleurochrysis carterae* (Q6RX25), and the stramenopiles *Cylindrotheca fusiformis* (Q6RX24) and *Mallomonas splendens* (Q9M7M6); parentheses show UniProt IDs. We aligned these sequences using MUSCLE (2) and used the 300-residue stretch described by Kiefel et al. (1) as an input for PSI-BLAST (3) against our nonredundant protein database at an *E*-value of 0.01. Of the resulting 12,963 sequences, 335 were eukaryotic. For maximum-likelihood analysis we retained 333 eukaryotic sequences (two low-quality sequences were removed), as well as 91 prokaryotic sequences representing α-proteobacteria (the closest relatives of mitochondria) and cyanobacteria (the closest relatives of chloroplasts). We constructed maximum-likelihood trees using RAxML (4). Numbers below the branches indicate bootstrap support from 100 trials; only values greater than 30% are shown. Solid triangles indicate collapsed clades; the number of collapsed sequences is given in brackets. Asterisks indicate eukaryotes known to contain bacterial endosymbionts. Plastid FtsZ proteins are related to cyanobaterial FtsZ and found in all plastid-containing lineages. Mitochondrial FtsZ proteins are related to α-proteobacterial FtsZ and are found only in amoebozoans, an apusozoan, red algae, glaucophytes, haptophytes, and stramenopiles. We find two homologs of mitochondrial FtsZ in the apusozoan *Thecamonas trahens*, a member of a sister group to the opisthokonts whose genome was recently sequenced (5). Both of these FtsZ genes are contiguous to intron-containing genes; one contains an intron itself, and another is predicted to localize to mitochondria [mitochondrial targeting potential 0.569; www.cbs.dtu.dk/services/TargetP (6)], suggesting they are bona fide mitochondrial FtsZs.

1. Kiefel BR, Gilson PR, Beech PL (2004) Diverse eukaryotes have retained mitochondrial homologues of the bacterial division protein FtsZ. *Protist* 155(1):105–115.
2. Edgar RC (2004) MUSCLE: Multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res* 32(5):1792–1797.
3. Altschul SF, et al. (1997) Gapped BLAST and PSI-BLAST: A new generation of protein database search programs. *Nucleic Acids Res* 25(17):3389–3402.
4. Stamatakis A (2014) RAxML version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* 30(9):1312–1313.
5. Torruella G, et al. (2012) Phylogenetic relationships within the Opisthokonta based on phylogenomic analyses of conserved single-copy protein domains. *Mol Biol Evol* 29(2):531–544.
6. Emanuelsson O, Brunak S, von Heijne G, Nielsen H (2007) Locating proteins in the cell using TargetP, SignalP and related tools. *Nat Protoc* 2(4):953–971.

Table S1. Functional annotation of dynamin variants

Class	Lineage/Species	Protein	ID	Signature	Functional annotation (ref.)
A	Alveolates/*Toxoplasma gondii*	DrpA	B9V0F0	A:A:AAA	Apicoplast fission (1)
A	Alveolates/*Tetrahymena thermophila*	Drp7	Q38KF4	::::::A:	MID (2)
A	Alveolates/*Plasmodium falciparum*	Dyn2	Q6KF55	::A::A::	Apicoplast fission (3)
A	Oomycete/*Phytophthora sojae*	VPS1	G4ZPT1	AAAAAAAA	Extracellular protein secretion (4)
A	Land plants/*Arabidopsis thaliana*	Drp3B	Q8LFT2	AAAAAA::	MID/PED (5, 6)
A	Land plants/*A. thaliana*	Drp3A	Q8S944	AAAAAA::	MID/PED (7, 8)
A	Red algae/*Cyanidioschyzon merolae*	Dnm1	Q84Y91	:A:::A::	MID/PED (9); Localized to cell surface (10)
A	Excavates/*Trypanosoma brucei*	Dlp1	Q8ITV0	AE:A::::	MID (11); VES (12)
A	Excavates/*Giardia lamblia*	Drp	E2RU04	::::::::	VES (13)
A	Excavates/*Trichomonas vaginalis*	DRP	A2FR27	A::A::::	Hydrogenosomal division (14)
A	Amoebozoans/*Entamoeba histolytica*	Dlp1	C4LX84	A:AA::::	Localized to nuclear membrane (15)
A	Amoebozoans/*Dictyostelium discoideum*	DymB	Q9U1M9	A:AA::::	PED (16)
A	Amoebozoans/*D. discoideum*	DymA	Q94464	AAAAAAAA	MID, VES (17)
A	Metazoans/*Homo sapiens*	Drp1	O00429	AAAAAAAA	MID/PED (18, 19)
A	Metazoans/*Caenorhabditis elegans*	Drp1	G5EDY8	AAAAAAAA	MID (20)
A	Fungi/*Schizosaccharomyces pombe*	Dnm1	Q09748	AAAAAAAA	MID/ PED (21, 22)
A	Fungi/*Saccharomyces cerevisiae*	Dnm1	P54861	AAAAAAAA	MID/ PED (23, 24)
A	Alveolates/*P. falciparum*	Dyn1	Q8IHR4	:U::::::	Localized to ER (25)
A	Alveolates/*T. thermophila*	Drp1	Q38KF7	:U::::::	VES (2)
A	Alveolates/*T. gondii*	DrpB	S8FEN4	:U::::::	VES (26)
A	Land plants/*A. thaliana*	Drp1A	P42697	AVAA:XYZ	PHR (27); VES (28, 29)
A	Land plants/*A. thaliana*	Drp1C	Q8LF21	AVAA:XYZ	PHR (30); Colocalized with clathrin (28)
A	Land plants/*A. thaliana*	Drp2B	Q9SE83	:W:-:XYZ	VES (29, 31); PHR (32)
A	Land plants/*Oryza sativa*	Drp2B	Q0DXR0	:W:-:XYZ	VES (33)
A	Metazoans/*Drosophila melanogaster*	Dyn1	P27619	AAA:::::	VES (34)
A	Metazoans/*H. sapiens*	Dyn1	Q05193	AAA:::::	VES (35, 36)
A	Metazoans/*H. sapiens*	Dyn2	P50570	AAA:::::	VES (36–38)
A	Fungi/*Candida albicans*	Vps1	Q5A1V4	AAASA:::	VES (39)
A	Fungi/*S. cerevisiae*	Vps1	P21576	AAASA:::	PED (24); VES (40); Vacuole fusion (41)
A	Fungi/*S. pombe*	Vps1	Q9URZ5	AAASA:::	PED (21); Vacuole tubulation (42)
B	Metazoans/*Haliotis discus*	Mx	A6MD73	BBBB:B:-	AVA (43)
B	Metazoans/*H. sapiens*	MxA	P20591	BBBB:BJK	AVA (44); Nuclear import (45)
B	Metazoans/*H. sapiens*	MxB	P20592	BBBB:BJK	Nuclear import (46)
B	Metazoans/*Mus musculus*	Mx1	P09922	BBBB:BJK	AVA (47); Nuclear localization (48)
B	Metazoans/*M. musculus*	Mx2	Q9WVP9	BBBB:BJK	AVA (49)
B	Metazoans/*Rattus norvegicus*	Mx1	P18588	BBBB:BJK	AVA (50)
B	Metazoans/*R. norvegicus*	Mx2	P18589	BBBB:BJK	AVA (50)
C	Land plants/*A. thaliana*	Drp5A	F4HPR5	LMN-----	CYT (51)
C	Amoebozoans/*D. discoideum*	DlpC	Q55AX0	LMN-----	CYT (51)
C	Land plants/*A. thaliana*	Drp5B	Q84N64	PQR-----	CHD (52); PED (53)
C	Land plants/*Physcomitrella patens*	Drp5B	B9WZC9	PQR-----	CHD (54)
	Metazoans/*H. sapiens*	Opa1	O60313		Mitochondrial fusion (55, 56)
	Fungi/*S. cerevisiae*	Mgm1	P32266		Mitochondrial fusion (57, 58)
	Fungi/*S. pombe*	Msp1	P87320		Mitochondrial fusion (59, 60)

Colors and abbreviations correspond to those used in Fig. 3. AVA, antiviral activity; CYT, cytokinesis; CHD, chloroplast division; MID, mitochondrial division; PED, peroxisomal division; PHR, phragmoplastin; VES, vesicle scission or secretion.

1. van Dooren GG, et al. (2009) A novel dynamin-related protein has been recruited for apicoplast fission in *Toxoplasma gondii*. *Curr Biol* 19(4):267–276.
2. Elde NC, Morgan G, Winey M, Sperling L, Turkewitz AP (2005) Elucidation of clathrin-mediated endocytosis in *tetrahymena* reveals an evolutionarily convergent recruitment of dynamin. *PLoS Genet* 1(5):e52.
3. Charneau S, et al. (2007) Characterization of PfDYN2, a dynamin-like protein of *Plasmodium falciparum* expressed in schizonts. *Microbes Infect* 9(7):797–805.
4. Li D, et al. (2013) PsVPS1, a dynamin-related protein, is involved in cyst germination and soybean infection of *Phytophthora sojae*. *PLoS ONE* 8(3):e58623.
5. Arimura S, Tsutsumi N (2002) A dynamin-like protein (ADL2b), rather than FtsZ, is involved in *Arabidopsis* mitochondrial division. *Proc Natl Acad Sci USA* 99(8):5727–5731.
6. Zhang X, Hu J (2009) Two small protein families, DYNAMIN-RELATED PROTEIN3 and FISSION1, are required for peroxisome fission in *Arabidopsis*. *Plant J* 57(1):146–159.
7. Arimura S, Aida GP, Fujimoto M, Nakazono M, Tsutsumi N (2004) *Arabidopsis* dynamin-like protein 2a (ADL2a), like ADL2b, is involved in plant mitochondrial division. *Plant Cell Physiol* 45(2):236–242.
8. Mano S, Nakamori C, Kondo M, Hayashi M, Nishimura M (2004) An *Arabidopsis* dynamin-related protein, DRP3A, controls both peroxisomal and mitochondrial division. *Plant J* 38(3): 487–498.
9. Imoto Y, et al. (2013) Single-membrane-bounded peroxisome division revealed by isolation of dynamin-based machinery. *Proc Natl Acad Sci USA* 110(23):9583–9588.
10. Nishida K, et al. (2003) Dynamic recruitment of dynamin for final mitochondrial severance in a primitive red alga. *Proc Natl Acad Sci USA* 100(4):2146–2151.
11. Morgan GW, Goulding D, Field MC (2004) The single dynamin-like protein of *Trypanosoma brucei* regulates mitochondrial division and is not required for endocytosis. *J Biol Chem* 279(11):10692–10701.
12. Chanez AL, Hehl AB, Engstler M, Schneider A (2006) Ablation of the single dynamin of *T. brucei* blocks mitochondrial fission and endocytosis and leads to a precise cytokinesis arrest. *J Cell Sci* 119(Pt 14):2968–2974.
13. Gaechter V, Schraner E, Wild P, Hehl AB (2008) The single dynamin family protein in the primitive protozoan *Giardia lamblia* is essential for stage conversion and endocytic transport. *Traffic* 9(1):57–71.
14. Wexler-Cohen Y, Stevens GC, Barnoy E, van der Bliek AM, Johnson PJ (2014) A dynamin-related protein contributes to *Trichomonas vaginalis* hydrogenosomal fission. *FASEB J* 28(3): 1113–1121.

15. Jain R, Shrimal S, Bhattacharya S, Bhattacharya A (2010) Identification and partial characterization of a dynamin-like protein, EhDLP1, from the protist parasite *Entamoeba histolytica*. *Eukaryot Cell* 9(1):215–223.
16. Rai A, Nöthe H, Tzvetkov N, Korenbaum E, Manstein DJ (2011) *Dictyostelium* dynamin B modulates cytoskeletal structures and membranous organelles. *Cell Mol Life Sci* 68(16):2751–2767.
17. Wienke DC, Knetsch ML, Neuhaus EM, Reedy MC, Manstein DJ (1999) Disruption of a dynamin homologue affects endocytosis, organelle morphology, and cytokinesis in *Dictyostelium discoideum*. *Mol Biol Cell* 10(1):225–243.
18. Smirnova E, Griparic L, Shurland DL, van der Bliek AM (2001) Dynamin-related protein Drp1 is required for mitochondrial division in mammalian cells. *Mol Biol Cell* 12(8):2245–2256.
19. Koch A, et al. (2003) Dynamin-like protein 1 is involved in peroxisomal fission. *J Biol Chem* 278(10):8597–8605.
20. Labrousse AM, Zappaterra MD, Rube DA, van der Bliek AM (1999) C. elegans dynamin-related protein DRP-1 controls severing of the mitochondrial outer membrane. *Mol Cell* 4(5): 815–826.
21. Jourdain I, Sontam D, Johnson C, Dillies C, Hyams JS (2008) Dynamin-dependent biogenesis, cell cycle regulation and mitochondrial association of peroxisomes in fission yeast. *Traffic* 9(3):353–365.
22. Jourdain I, Gachet Y, Hyams JS (2009) The dynamin related protein Dnm1 fragments mitochondria in a microtubule-dependent manner during the fission yeast cell cycle. *Cell Motil Cytoskeleton* 66(8):509–523.
23. Bleazard W, et al. (1999) The dynamin-related GTPase Dnm1 regulates mitochondrial fission in yeast. *Nat Cell Biol* 1(5):298–304.
24. Kuravi K, et al. (2006) Dynamin-related proteins Vps1p and Dnm1p control peroxisome abundance in *Saccharomyces cerevisiae*. *J Cell Sci* 119(Pt 19):3994–4001.
25. Li H, et al. (2004) Isolation and functional characterization of a dynamin-like gene from *Plasmodium falciparum*. *Biochem Biophys Res Commun* 320(3):664–671.
26. Breinich MS, et al. (2009) A dynamin is required for the biogenesis of secretory organelles in *Toxoplasma gondii*. *Curr Biol* 19(4):277–286.
27. Kang BH, Busse JS, Bednarek SY (2003a) Members of the Arabidopsis dynamin-like gene family, ADL1, are essential for plant cytokinesis and polarized cell growth. *Plant Cell* 15(4):899–913.
28. Konopka CA, Backues SK, Bednarek SY (2008) Dynamics of *Arabidopsis* dynamin-related protein 1C and a clathrin light chain at the plasma membrane. *Plant Cell* 20(5):1363–1380.
29. Fujimoto M, et al. (2010) *Arabidopsis* dynamin-related proteins DRP2B and DRP1A participate together in clathrin-coated vesicle formation during endocytosis. *Proc Natl Acad Sci USA* 107(13):6094–6099.
30. Kang BH, Rancour DM, Bednarek SY (2003b) The dynamin-like protein ADL1C is essential for plasma membrane maintenance during pollen maturation. *Plant J* 35(1):1–15.
31. Jin JB, et al. (2001) A new dynamin-like protein, ADL6, is involved in trafficking from the trans-Golgi network to the central vacuole in *Arabidopsis*. *Plant Cell* 13(7):1511–1526.
32. Fujimoto M, Arimura S, Nakazono M, Tsutsumi N (2008) *Arabidopsis* dynamin-related protein DRP2B is co-localized with DRP1A on the leading edge of the forming cell plate. *Plant Cell Rep* 27(10):1581–1586.
33. Xiong G, et al. (2010) The rice dynamin-related protein DRP2B mediates membrane trafficking, and thereby plays a critical role in secondary cell wall cellulose biosynthesis. *Plant J* 64(1):56–70.
34. van der Bliek AM, Meyerowitz EM (1991) Dynamin-like protein encoded by the *Drosophila* shibire gene associated with vesicular traffic. *Nature* 351(6325):411–414.
35. van der Bliek AM, et al. (1993) Mutations in human dynamin block an intermediate stage in coated vesicle formation. *J Cell Biol* 122(3):553–563.
36. Altschuler Y, et al. (1998) Redundant and distinct functions for dynamin-1 and dynamin-2 isoforms. *J Cell Biol* 143(7):1871–1881.
37. Kreitzer G, Marmorstein A, Okamoto P, Vallee R, Rodriguez-Boulan E (2000) Kinesin and dynamin are required for post-Golgi transport of a plasma-membrane protein. *Nat Cell Biol* 2(2):125–127.
38. Henley JR, Krueger EW, Oswald BJ, McNiven MA (1998) Dynamin-mediated internalization of caveolae. *J Cell Biol* 141(1):85–99.
39. Bernardo SM, Khalique Z, Kot J, Jones JK, Lee SA (2008) *Candida albicans* VPS1 contributes to protease secretion, filamentation, and biofilm formation. *Fungal Genet Biol* 45(6):861–877.
40. Smaczynska-de Rooij II, et al. (2010) A role for the dynamin-like protein Vps1 during endocytosis in yeast. *J Cell Sci* 123(Pt 20):3496–3506.
41. Alpadi K, et al. (2013) Dynamin-SNARE interactions control trans-SNARE formation in intracellular membrane fusion. *Nat Commun* 4:1704.
42. Röthlisberger S, Jourdain I, Johnson C, Takegawa K, Hyams JS (2009) The dynamin-related protein Vps1 regulates vacuole fission, fusion and tubulation in the fission yeast, *Schizosaccharomyces pombe*. *Fungal Genet Biol* 46(12):927–935.
43. De Zoysa M, et al. (2007) First report of invertebrate Mx: Cloning, characterization and expression analysis of Mx cDNA in disk abalone (*Haliotis discus discus*). *Fish Shellfish Immunol* 23 (1):86–96.
44. Pavlovic J, Zürcher T, Haller O, Staeheli P (1990) Resistance to influenza virus and vesicular stomatitis virus conferred by expression of human MxA protein. *J Virol* 64(7):3370–3375.
45. Kochs G, Haller O (1999) Interferon-induced human MxA GTPase blocks nuclear import of Thogoto virus nucleocapsids. *Proc Natl Acad Sci USA* 96(5):2082–2086.
46. King MC, Raposo G, Lemmon MA (2004) Inhibition of nuclear import and cell-cycle progression by mutated forms of the dynamin-like GTPase MxB. *Proc Natl Acad Sci USA* 101(24): 8957–8962.
47. Horisberger MA, Staeheli P, Haller O (1983) Interferon induces a unique protein in mouse cells bearing a gene for resistance to influenza virus. *Proc Natl Acad Sci USA* 80(7):1910–1914.
48. Haller O, Kochs G (2002) Interferon-induced mx proteins: Dynamin-like GTPases with antiviral activity. *Traffic* 3(10):710–717.
49. Zürcher T, Pavlovic J, Staeheli P (1992) Mouse Mx2 protein inhibits vesicular stomatitis virus but not influenza virus. *Virology* 187(2):796–800.
50. Meier E, Kunz G, Haller O, Arnheiter H (1990) Activity of rat Mx proteins against a rhabdovirus. *J Virol* 64(12):6263–6269.
51. Miyagishima SY, Kuwayama H, Urushihara H, Nakanishi H (2008) Evolutionary linkage between eukaryotic cytokinesis and chloroplast division by dynamin proteins. *Proc Natl Acad Sci USA* 105(39):15202–15207.
52. Gao H, Kadirjan-Kalbach D, Froehlich JE, Osteryoung KW (2003) ARC5, a cytosolic dynamin-like protein from plants, is part of the chloroplast division machinery. *Proc Natl Acad Sci USA* 100(7):4328–4333.
53. Zhang X, Hu J (2010) The *Arabidopsis* chloroplast division protein DYNAMIN-RELATED PROTEIN5B also mediates peroxisome division. *Plant Cell* 22(2):431–442.
54. Sakaguchi E, et al. (2011) Three dynamin-related protein 5B genes are related to plastid division in *Physcomitrella patens*. *Plant Sci* 180(6):789–795.
55. Olichon A, et al. (2002) The human dynamin-related protein OPA1 is anchored to the mitochondrial inner membrane facing the inter-membrane space. *FEBS Lett* 523(1–3):171–176.
56. Cipolat S, Martins de Brito O, Dal Zilio B, Scorrano L (2004) OPA1 requires mitofusin 1 to promote mitochondrial fusion. *Proc Natl Acad Sci USA* 101(45):15927–15932.
57. Wong ED, et al. (2003) The intramitochondrial dynamin-related GTPase, Mgm1p, is a component of a protein complex that mediates mitochondrial fusion. *J Cell Biol* 160(3):303–311.
58. Sesaki H, Southard SM, Yaffe MP, Jensen RE (2003) Mgm1p, a dynamin-related GTPase, is essential for fusion of the mitochondrial outer membrane. *Mol Biol Cell* 14(6):2342–2356.
59. Pelloquin L, Belenguer P, Menon Y, Gas N, Ducommun B (1999) Fission yeast Msp1 is a mitochondrial dynamin-related protein. *J Cell Sci* 112(Pt 22):4151–4161.
60. Guillou E, Bousquet C, Daloyau M, Emorine LJ, Belenguer P (2005) Msp1p is an intermembrane space dynamin-related protein that mediates mitochondrial fusion in a Dnm1p-dependent manner in *S. pombe*. *FEBS Lett* 579(5):1109–1116.

Dataset S1. Complete dynamin dataset, cluster attributes, and protist genome databases

Dataset S1

Sheet 1: Complete dynamin dataset. Columns: (*A*) UniProt ID or protist proteome ID. (*B*) Classes "A", "B", and "C" indicate major functional superclasses. "Mgm1" and "OPA1" are mitochondrial fusion dynamins of fungi and metazoans, respectively. The other lineage-specific clusters are designated "C (Alveolate specific)" and "Bacterial". (*C*) Species. (*D*) Group. (*E*) Subgroup. (*F*) Domain architecture. (*G*) List of splice isoforms. (*H*) Short eight-character signature. Clusters that span supergroups or other relevant lineages are indicated by a segment-wise letter; most supergroup-specific clusters or unclustered sequences are indicated by a colon (:); missing segments are indicated by a dash (-). (*I–P*) Numerical cluster labels in the format [segment.label]. (*Q–X*) Segment coordinates according to Pfam HMM profiles (ND, PF00350; MD, PF01031; GED, PF02212). (*Y–AF*) Segment coordinates along the protein sequence. If a protein has repeated segments, all occurrences of the segment are shown. Sheet 2: Cluster attributes. Columns: (*A*) Segment/superclass. (*B*) Cluster label. (*C*) BLAST *P* value at which each cluster is called in CLANS. (*D*) Number of sequences. (*E*) Attribute 1: Median intracluster percentage of identity. (*F*) Attribute 2: Root species (for clusters containing proteins from multiple supergroups). LCA_Am_Ar, last common ancestor of amoebozoans and archaeplastids; LCA_Ar_St, last common ancestor of archaeplastids and stramenopiles; LECA, last eukaryotic common ancestor. (*G*) Monophyly bootstrap support, number of trees out of 1,000 (for clusters with 10 or more members). Sheet 3: Protist genome databases. Shown are species, taxonomic IDs, and source databases for protist genomes and proteomes used to construct the nonredundant protein dataset.

Dataset S2. FASTA alignment of dynamin sequences

Dataset S2

The alignment of 3,841 sequences is generated by concatenating the 556 residues corresponding to Pfam HMMs for the ND (PF00350, 168 residues), MD (PF01031, 296 residues), and GED (PF02212, 92 residues). Protein IDs are those used in Dataset S1.