nature COMMUNICATIONS

ARTICLE

Received 1 Apr 2016 | Accepted 5 Aug 2016 | Published 15 Sep 2016

DOI: 10.1038/ncomms12847 OPEN

An experimental phylogeny to benchmark ancestral sequence reconstruction

Ryan N. Randall[1], Caelan E. Radford[1], Kelsey A. Roof[1], Divya K. Natarajan[1] & Eric A. Gaucher[1,2]

Ancestral sequence reconstruction (ASR) is a still-burgeoning method that has revealed many key mechanisms of molecular evolution. One criticism of the approach is an inability to validate its algorithms within a biological context as opposed to a computer simulation. Here we build an experimental phylogeny using the gene of a single red fluorescent protein to address this criticism. The evolved phylogeny consists of 19 operational taxonomic units (leaves) and 17 ancestral bifurcations (nodes) that display a wide variety of fluorescent phenotypes. The 19 leaves then serve as 'modern' sequences that we subject to ASR analyses using various algorithms and to benchmark against the known ancestral genotypes and ancestral phenotypes. We confirm computer simulations that show all algorithms infer ancient sequences with high accuracy, yet we also reveal wide variation in the phenotypes encoded by incorrectly inferred sequences. Specifically, Bayesian methods incorporating rate variation significantly outperform the maximum parsimony criterion in phenotypic accuracy. Subsampling of extant sequences had minor effect on the inference of ancestral sequences.

[1] School of Biological Sciences, Georgia Institute of Technology, Atlanta, Georgia 30332, USA. [2] Institute for Bioengineering and Biosciences, Georgia Institute of Technology, Atlanta, Georgia 30332, USA. Correspondence and requests for materials should be addressed to E.A.G. (email: eric.gaucher@biology.gatech.edu).

Ancestral sequence reconstruction (ASR) is the process of analyzing modern sequences within an evolutionary/phylogenetic context to infer the ancestral sequences at particular nodes of a tree[1]. These ancient sequences are most often then synthesized, recombinantly expressed in laboratory microorganisms or cell lines, and then characterized to reveal the ancient properties of the extinct biomolecules[2–6]. This process has produced tremendous insights into the mechanisms of molecular adaptation and functional divergence[7]. Despite such insights, a major criticism of ASR is the general inability to benchmark accuracy of the implemented algorithms. It is difficult to benchmark ASR for many reasons. Notably, genetic material is not preserved in fossils on a long enough time scale to satisfy most ASR studies (many millions to billions of years ago), and it is not yet physically possible to travel back in time to collect samples.

To overcome these limitations, we exploited an under-utilized yet effective procedure to develop a phylogeny in the laboratory[8]. The benefits of the procedure are at least twofold: (1) we can accelerate the process of evolution that generates the vertical inheritance of genetic information necessary for the functional divergence of encoded phenotypes and (2) we have a known record of the ancestral genotypes and phenotypes throughout the experimental phylogeny. The goal of the phylogeny is thus to create an opportunity to evolve sequences within a controlled framework that adds biological reality given practical limitations. We elected to build the phylogeny using a single monomeric red fluorescent protein (FP), since it is known that FP colour phenotypes are readily modified by a tractable number of amino acid replacements[9,10]. The experimental phylogeny then provides us with an opportunity to benchmark the performance of algorithms that infer ancient sequences. In particular, we were interested in determining the accuracy of algorithms when inferring ancestral phenotypes since computer simulations have shown that these algorithms infer ancient genotypes with reasonably high accuracy[11–13]. Our benchmarking exercise focused on Bayesian versus maximum parsimony (MP) algorithms, the effect of rate variation when modelled as a discrete gamma distribution[14], subsamples of taxa to infer ancestral sequences, and species-tree-aware versus unaware approaches within the Bayesian framework[15,16].

Our study confirms that all ASR algorithms correctly infer the vast majority of residues in ancestral sequences. Yet, these algorithms differ in the amino acid identities of the small number of sites that are incorrectly inferred. Here we demonstrate that these incorrectly inferred residues can indeed influence the protein phenotypes of the encoded ancestral sequences and that various parameters incorporated into evolutionary models affect these incorrectly inferred sites.

Results

Evolving the experimental phylogeny. We built the FP phylogeny from a single gene using random mutagenesis PCR (Fig. 1). Each round of PCR produced numerous variants, or descendants, of which only one was retained for the next round of random mutagenesis, unless a bifurcation was being incorporated

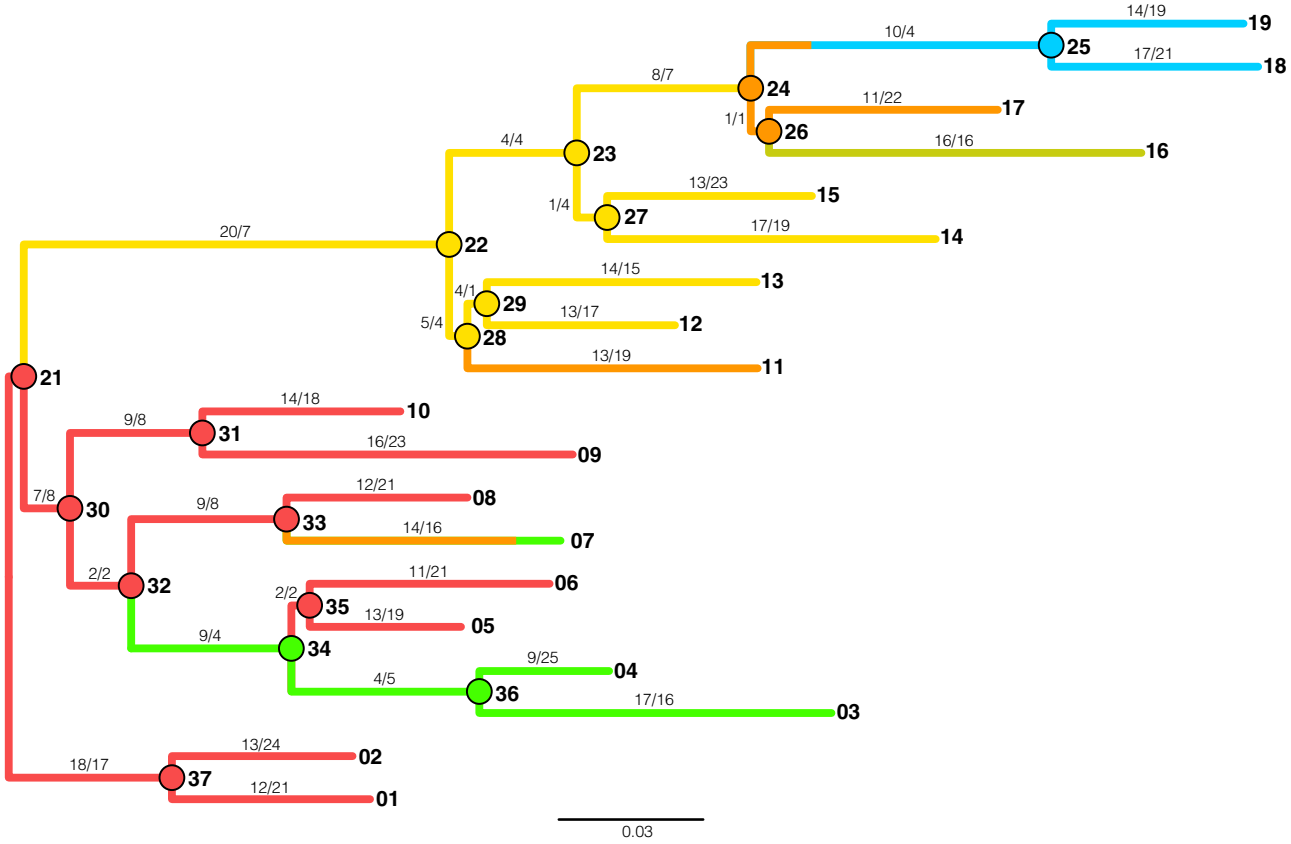


Figure 1 | Phylogram of the experimental phylogeny initiated from a single red FP gene. Scale bar represents amino acid replacements per site per unit evolutionary time. The colour of each branch reflects the colour-class phenotype (emission) of the node protein for internal branches or the leaf protein for tip branches (except for the branch connecting node 33 to leaf 7 that transitions through an orange intermediate). Nodes and tips are numbered for reference. Nonsynonymous and synonymous substitutions are shown along each branch, respectively. The experiment began near node 21 with a single red FP gene and proceeded by random-mutagenesis PCR.

into the tree, in which case two variants would be allowed to progress (Supplementary Fig. 1). To best create biological context, the branch lengths, the number of synonymous and nonsynonymous substitutions, base frequencies, and phenotypic diversity followed that of natural FP sequences[17]. In total, the phylogeny contains 19 operational taxonomic units (leaves of the tree) that serve as 'modern' sequences and 17 ancestral bifurcations (nodes of the tree) that serve as true or known 'ancient' sequences. The phylogeny contains a total of 833 mutations (461 synonymous and 372 nonsynonymous), with a small percentage of these experiencing homoplasy, but no insertion/deletion events (Supplementary Table 1). Transitions were more abundant than transversions, 64% versus 36%, respectively. Most branches were evolved under purifying selection except when selecting for modifications in the phenotypic emission properties of the FPs. Colour properties of the evolved proteins included variations of red, orange, yellow, green and blue (Supplementary Fig. 2). The order and distribution of colour emission phenotypes for FP proteins were mapped onto the phylogeny (Supplementary Fig. 3).

Inferring ancient sequences. The 19 leaf-sequences were collected and subjected to ASR analyses. The sequences were analysed using MP and Bayesian algorithms, and for the Bayesian approach[18,19], we analysed the effects of incorporating rate variation (gamma distribution [Γ] versus no gamma distribution, or rate heterogeneity versus rate homogeneity) and the effect of accounting for possible gene duplication, horizontal transfer, or gene loss events (so-called species-tree-aware trees, as implemented in PhyloBayes)[16]. Figure 2 shows the results from five different ASR analyses across all nodes of the phylogeny as a function of the number of incorrectly inferred ancestral amino acids. This figure shows the expected pattern that all ASR procedures perform well for more derived nodes, while all procedures perform worse for more basal nodes. In terms of raw percentage of correctly inferred residues, most procedures recapitulated reality (Supplementary Table 2). The Bayesian approaches that incorporated rate variation using a species-tree-unaware tree were the most accurate (PAML_Γ and FastML_Γ, Supplementary Table 3a,c), then Bayesian without rate variation (PAML), followed by MP, and finally by Bayesian with rate variation and species-tree-aware tree using PhyloBayes (PHYLO_Γ). PAML and FastML were expected to perform analogously since they are similar implementations of the Bayesian algorithm. Total accuracy for the five procedures ranged between 97.88 and 98.17% (Supplementary Table 2), thus reflecting the general sequence accuracy of ASR algorithms.

Characterizing the evolved and inferred protein phenotypes. Despite the overall sequence accuracies of the five procedures, we questioned whether the phenotypes associated with the incorrectly inferred ancestral sequences are themselves incorrect. We synthesized, expressed and purified each incorrectly inferred ancestral protein at each node of the tree for each procedure to determine whether there was variation in the phenotypes for the incorrect proteins compared with the true ancestral phenotypes. These 34 proteins were phenotypically characterized in terms of their extinction coefficients (ε), quantum yield (Φ) and brightness (product of ε and Φ) (Supplementary Table 3a,b). Properties of the resurrected ancestral proteins were compared with the true ancestral proteins to determine the percent error in phenotypes. Figure 3 shows that significant variation in phenotypic error exists between the five procedures. PAML_Γ and FastML_Γ had significantly less error than PAML without rate variation and PHYLO_Γ at the 95% confidence interval, and less error than MP at the 99% level when characterizing extinction coefficients. Conversely, PHYLO_Γ generated significantly lower error

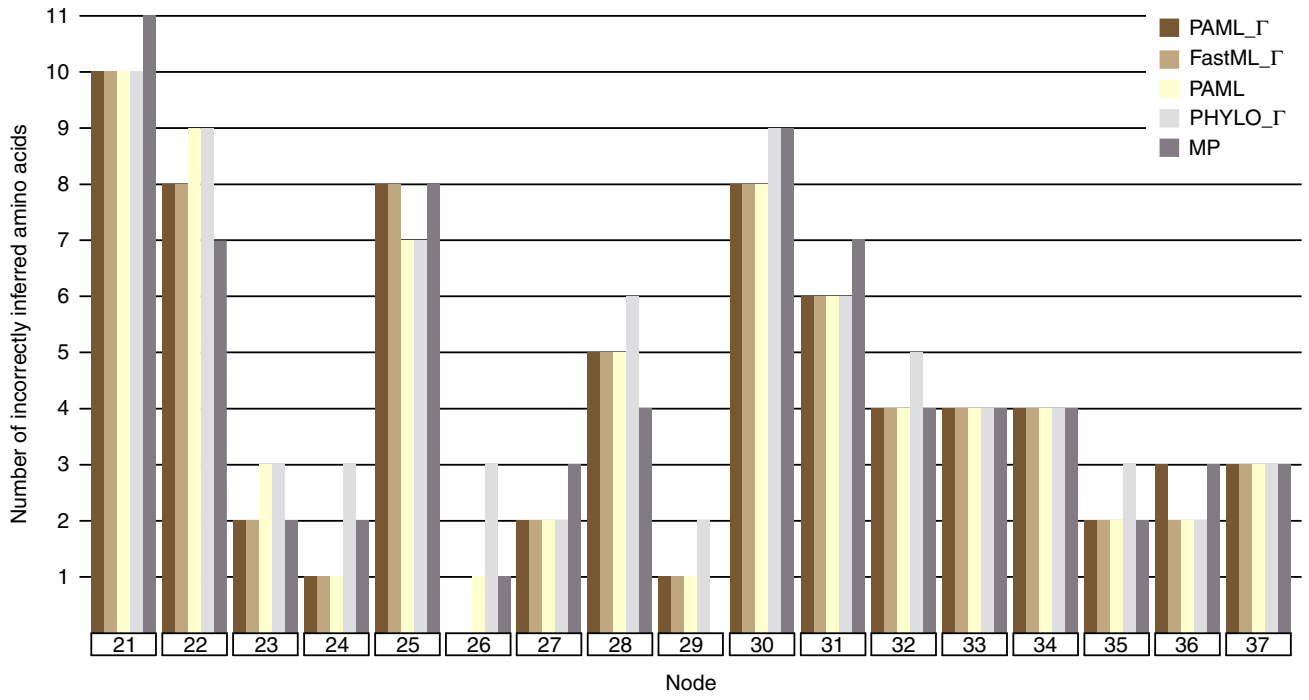


Figure 2 | Number of incorrectly inferred amino acid sites for each node of the phylogeny. The 19 leaf sequences from Fig. 1 were subjected to ASR analyses using Bayesian (PAML, FastML, PhyloBayes) with or without rate variation modelled as a gamma distribution (Γ), as well as parsimony (MP). The inferred sequences were then compared to the true ancestral sequences from the 17 ancestral nodes in Fig. 1. Dark brown bars are PAML with a gamma distribution, light brown bars are FastML with a gamma distribution, yellow bars are PAML without gamma, light grey bars are PhyloBayes with a gamma distribution, and dark grey bars are maximum parsimony. Colour code is irrespective of FP colour emission phenotype.

compared with MP for quantum yield at the 95% level, with the other three Bayesian procedures only slightly worse than PHYLO_Γ. The brightness phenotype demonstrated that PAML_Γ and FastML_Γ carried over their significantly lower error compared to MP at the 99% level, and that PAML without rate variation and PHYLO_Γ displayed similar amounts of error less than MP. None of the five procedures displayed significant error for emission wavelengths, and as seen in Supplementary Table 3a, little error was displayed for excitation wavelengths.

Discussion

We have applied brute-force random-mutagenesis and guided-selection to generate an experimental phylogeny of synthetic FPs to recapitulate evolutionary processes that govern natural FPs[17]. This phylogeny contained ample phenotypic diversity analogous to most gene families subjected to ASR studies. Despite best efforts, we anticipate that experimental bias exists within the phylogeny to some degree, but we should not be paralysed since natural molecular systems routinely display bias (for example, G + C, transitions, base composition and so on) and parameter-based models account for such bias. The experimental phylogeny allowed us to verify that ASR often generates correct ancestral phenotypes even when the wrong ancestral sequences have been inferred in our FP system (Supplementary Table 3a,c). We anticipate that such accuracy would hold true for other laboratory-evolved gene families since there is no reason to think that FP evolution involves an extraordinary mechanism *per se*. However, improvements can still be made to these phylogenetic algorithms. In particular, we tested known limitations of these algorithms by purposely invoking homoplasy into the experimental phylogeny. The five procedures performed generally well with reversions, and parallel and convergent amino acid replacements, as long as they occurred along sufficiently long branches (Supplementary Fig. 1).

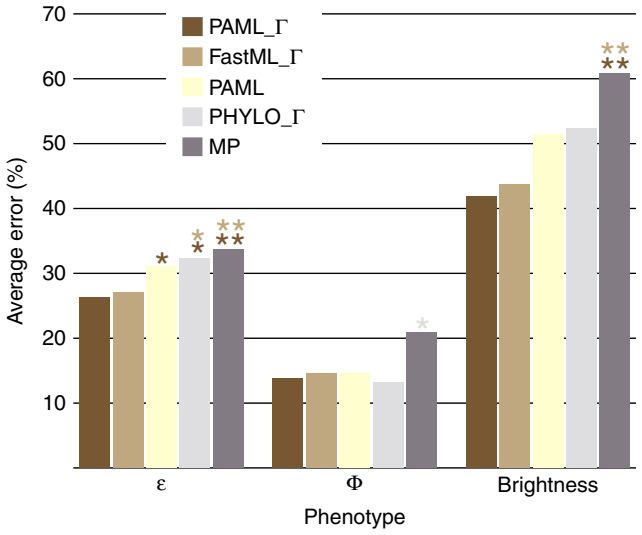


Figure 3 | Average phenotypic error across all nodes for the five ASR procedures. Extinction coefficient (ε), quantum yield (Φ), and brightness (product of ε and Φ) were determined for all incorrectly inferred ancestral FP proteins and compared to the properties of the true ancestral protein at each node and reported as a function of percent error. Dark brown bars are PAML with a gamma distribution, light brown bars are FastML with a gamma distribution, yellow bars are PAML, light grey bars are PhyloBayes with a gamma distribution, and dark grey bars are maximum parsimony. Single and double asterisks represent confidence at 95% and 99% levels, respectively, and are coloured according to the respective procedure that has significantly less error.

This was not the case for one scenario though. Ancestral node 32 (An32) accumulates an amino acid replacement (Y120C) that causes the colour phenotype to switch from red to green by An34 (Supplementary Fig. 1). The reversion of C120Y then occurs along the short branch giving rise to a red An35. None of the five procedures could account for this mode of homoplasy, thus they all predicted that An34 was red. Notably, this is the only incorrectly inferred protein displaying emission in a separate colour class than the true ancestor. Interestingly, all five procedures predicted sequences that encode the correct colour emission for An35 yet all of these sequences encode the highly incorrect quantum yield carried over from An34 (Supplementary Table 3a). This result demonstrates that protein properties encoded by incorrect sequences can propagate throughout nodes connected by short branches—a caution for ASR studies. Notwithstanding such known difficulties of homoplasy, all five procedures performed admirably despite the phylogeny experiencing dramatic phenotypic changes. For instance, the branch connecting An33 and leaf 7 switched from red to orange to green colour phenotypes, but all five procedures correctly inferred the colour phenotype despite each incorrectly inferring four amino acid sites (Figs 1 and 2, Supplementary Table 3a). Similarly, all five procedures correctly inferred the colour phenotype of An24 despite that only one of the four descendent leaves had the same colour phenotype as the ancestor (and despite that the five procedures incorrectly inferred 1–3 amino acid sites; Figs 1 and 2, Supplementary Table 3a).

Taxon sampling in phylogenetics has been greatly debated over the past 20 years[20]. Central to this debate is whether sub-samplings of taxa lead to inferences of incorrect phylogenies. Computer simulations and large data sets of angiosperms have supported the conclusion that robust taxa samples are superior than smaller subsamples of taxa[20,21], while others have argued against this conclusion[22,23]. To address the issue of sub-sampling of taxa in the inference of ancestral sequences (but not the topology itself), we generated two diverse subsamples from our 19 leaf sequences. Our analysis focused on the last common ancestor (most ancient divergence, An21 from Fig. 1) of the phylogeny since this sequence is theoretically the most difficult to correctly infer. One subsample incorporated every other sequence along the continuum of leaf sequences, while the other subsample consisted of closely related groups of leaf sequences (Supplementary Fig. 4). Both of these subsamples utilized half the number of leaf sequences and inferred a sequence for the last common ancestor of the two reduced phylogenies. Each of the two ancestors differed at only a single position when compared with the 10 incorrect residues inferred using the entire 19 sequences with WAG_Γ (at An21). The single amino acid position that differed from each subsample analysis experienced sufficient homoplasy throughout the assembly of the experimental phylogeny yet was not associated with phenotypic change. Although these sub-sampling analyses focus on the effects of using half of the leaf sequences to infer the last common ancestor, and showed little effect, additional sub-samplings should be explored to gain additional insights into the effects of subsamples in ASR studies[20].

Overall, our experimental phylogeny has allowed ASR algorithms to be benchmarked for the first time against biologically encoded sequences. Our analyses focused on amino acid reconstructions since they performed slightly worse than DNA- and codon-based analyses (98.4% and 98.3% of sites correctly inferred, respectively, Supplementary Table 2). These analyses confirm computational predictions regarding the accuracy of ASR but extend our understanding of the algorithms by revealing the phenotypes associated with incorrectly inferred ancestral sequences. All of the tested ASR algorithms and

procedures work generally well in terms of capturing the true ancestral phenotype even when the true ancestral genotype is not fully recapitulated. This finding should give the ASR field confidence that ancestral phenotypes are encoded correctly even if some residues are incorrectly inferred—assuming such sites do not drive phenotypes. Our experimental phylogeny has also allowed us to determine nuances of ASR analyses. For instance, incorporating rate variation using a discrete gamma distribution had little effect on the total number of incorrectly inferred amino acids (71 with gamma versus 72 without gamma, Supplementary Table 2), however, the positions of these incorrectly inferred sites differed, and more interestingly, the encoded phenotypes differed substantially in terms of brightness error between the two procedures (42% with gamma versus 51% without gamma, Supplementary Table 3c). Intriguingly, our analyses did not find a strong correlation between the number of incorrectly inferred residues versus the errors in the measured phenotypes. Since more-ancient nodes often contain more incorrectly inferred residues[11,13], this suggests that more-ancient nodes are not necessarily encoding furtherly biased phenotypes. Further, our analyses demonstrated that incorporating a species-tree-aware procedure had the same overall effect as not incorporating rate variation. This is not to say that species-tree-aware procedures inherently mislead, rather, our experimental phylogeny was void of gene loss/gain/duplication so a PhyloBayes analysis would be over-parameterized for our dataset. But our results do suggest that a species-tree-unaware procedure is more appropriate in the absence of gene loss/gain/duplication or lineage sorting. Finally, our analyses demonstrate that Bayesian procedures produce more accurate ancestral phenotypes than the MP criterion, regardless of the Bayesian parameters tested. This is not a general abomination against parsimony, as it performed quite accurately for many nodes in the phylogeny. Rather, if the goal is to generate the most accurate ancestral phenotypes possible, then ASR studies on gene families that have evolved (at least) analogously to our laboratory FP system would benefit most from Bayesian procedures.

We anticipate that providing assurance on the accuracy of ASR will allow the field to move forward in novel ways, such as the synthesis of complete ancestral genomes[24–26], development of mechanistic models of protein evolution[27,28], contribute to the debate about alternative ancestral sequences[11,13,18] and to support synthetic biology approaches that exploit ASR for applied purposes[29–31].

Methods

Evolving the experimental phylogeny. The following two primers were used for PCR mutagenesis: FP Random Forward (5′-CTGGTCGGCCATATGGCGTCTT CTGAAGACGTTATC-3′) and FP Random Reverse (5′-CGGATCCTCGAGCTA TTACGCACCGGTAGAGTG-3′). Random mutagenesis of *mRFP1* was performed using the GeneMorph II Random Mutagenesis Kit (Stratagene). Each reaction was performed in 50 µl and consisted of the following: 425–625 ng template plasmid, 0.25 µl forward primer, 0.25 µl reverse primer, 1 µl of 40 mM dNTP stock, 5 µl 10 × Mutazyme II reaction buffer, 1 µl Mutazyme II DNA polymerase. PCR was performed using the following conditions: initial incubation at 95 °C 2 min then 95 °C 30 s, 59 °C 30 s, 72 °C 1 min, repeat × 29, final incubation at 72 °C 10 min. PCR products were purified using Qiagen PCR clean up kit following the manufacturer's protocol. Purified mutagenesis PCR DNA was digested in a 50 µl reaction at 37 °C between 16–48 h and included the following: FP DNA, 1 µl *Xho*I, 1 µl *Nde*I, 5 µl Buffer 4, 0.5 µl BSA. Digested DNA was cleaned up using Qiagen's PCR clean up kit following the manufacture's protocol. The digested and cleaned FP mutant genes were ligated into pET-15b (Novagen) according to the following protocol: 100 ng digested pET-15b Vector, 20 ng digested FP gene, 0.5 µl T4 Ligase, 1–2 µl 10 × T4 Ligase Buffer in a 10–20 µl reaction. Plasmids containing mutated mRFP 1.0 were transformed into expression host *E. coli* BL21(DE3). Eight to twenty colonies expressing FP genes were selected and sequenced (GeneWiz). Sequence data were analysed using CLC Bio software version 4.1.2. The average PCR FP variant contained 1–4 base substitutions per round of random mutagenesis given the conditions above (conditions optimized for this mutation load). One mutant was retained after each round of mutagenesis and used for subsequent rounds of random mutagenesis. Mutants were selected to balance the frequency of synonymous and nonsynonymous substitutions along branches of the experimental gene phylogeny. In some instances, two mutants were retained after a round of mutagenesis to bifurcate the phylogeny. The experimental phylogeny was initiated by replacing seven amino acid positions known to affect emission phenotype[32] (Supplementary Fig. 1), all other mutations were random. All leaf and internal node amino acid sequences are proved (Supplementary Note 1).

Ancestral sequence reconstruction. The 19 'modern' sequences at the tips (leaves) of the FP phylogeny were used to computationally reconstruct infer ancestral sequences at all internal nodes of the tree using the evolved (known) topology. Marginally reconstructed ancestral sequences were inferred using Bayesian approaches that incorporated the WAG amino acid replacement matrix (PAML, FastML and PhyloBayes [CAT]), with or without rate variation as modelled by a discrete gamma distribution (four rate categories), and ancestral sequences were also inferred with the MP criterion (as implemented in PAML). DNA and codon-based analyses were performed only in PAML using HKY85 + GC and M0(F3x4), respectively. ProtTest v3.2 was used to analyse the various models according to the AIC criterion (AIC weight was 100% for WAG_Γ)[33].

Protein expression and purification. FP variants were transformed into BL21(DE3) bacterial cells. Transformed cells grew overnight at 37 °C on LB-agar supplemented with 50 µg ml^{-1} carbenicillin. A single colony was inoculated and grown overnight in 5 ml LB/carbenicillin. The next day, the culture was used to inoculate LB/carbenicillin using a 3:100 ratio. Cells were induced at OD_{600} of 0.55–0.9 with isopropyl β-D-1-thiogalactopyranoside (IPTG) to a final concentration of 100 µM. Cultures continued to grow for 4–6 h at 37 °C and were then harvested by centrifugation and stored at − 80 °C. Cell pellets were lysed with BugBuster (Novagen) and purified by Ni-NTA agarose (Qiagen) in elution buffer containing 500 mM imidazole per manufacturer's protocol. Imidazole was removed from the protein samples via buffer exchange using 20-kDa concentrators (Pierce) or dialysis into 50 mM Tris buffer pH 7.5. Purity of protein sample was assessed by SDS–PAGE.

Protein characterization and spectroscopic studies. Absorption spectra of purified protein were recorded on a Varian Cary 50 ultraviolet–visible spectrophotometer (Supplementary Fig. 5). Excitation and emission spectra were recorded on a Varian Cary Eclipse fluorescence spectrophotometer. All measurements were performed in quartz cuvettes at ambient temperature and purified protein sample was diluted in 5 mM Tris/HCl (pH 7.5). Quantum yield experiments were performed as described[34,35] and variant proteins were compared to equally absorbing solutions of mRFP1 for red emitting variants, TagGFP2 for yellow and orange emitting variants, or TagBFP for green and blue emitting variants. Quantum yield values were computed using a|e UV–vis-IR Spectral software. Extinction coefficients were determined as described[10]. Molar extinction coefficient values were calculated using the maximum absorbance at wavelength of maximum excitation.

The relative errors between the inferences across all nodes for each procedure versus the true ancestors were determined for extinction coefficient, quantum yield and brightness. Bootstrap analysis were performed; the actual sum of the differences in relative error for each node between two methods was first determined. Then, a distribution of sums of differences in relative error for each node between any two procedures was made for 100 bootstrap replicate data sets generated by assigning error values from the appropriate nodes of the original data set with replacement. The percentile of the actual sum of differences in relative error in the distribution was determined and reported.

Data availability. All relevant data are available from the authors by request.

References

1. Thornton, J. W. Resurrecting ancient genes: experimental analysis of extinct molecules. *Nat. Rev. Genet.* **5,** 366–375 (2004).
2. Anderson, D. P. *et al.* Evolution of an ancient protein function involved in organized multicellularity in animals. *Elife* **5,** e10147 (2016).
3. Bickelmann, C. *et al.* The molecular origin and evolution of dim-light vision in mammals. *Evolution* **69,** 2995–3003 (2015).
4. Hobbs, J. K., Prentice, E. J., Groussin, M. & Arcus, V. L. Reconstructed ancestral enzymes impose a fitness cost upon modern bacteria despite exhibiting favourable biochemical properties. *J. Mol. Evol.* **81,** 110–120 (2015).
5. Kratzer, J. T. *et al.* Evolutionary history and metabolic insights of ancient mammalian uricases. *Proc. Natl Acad. Sci. USA* **111,** 3763–3768 (2014).
6. Wilson, C. *et al.* Kinase dynamics. Using ancient protein kinases to unravel a modern cancer drug's mechanism. *Science* **347,** 882–886 (2015).
7. Dean, A. M. & Thornton, J. W. Mechanistic approaches to the study of evolution: the functional synthesis. *Nat. Rev. Genet.* **8,** 675–688 (2007).
8. Hillis, D. M., Bull, J. J., White, M. E., Badgett, M. R. & Molineux, I. J. Experimental phylogenetics: generation of a known phylogeny. *Science* **255,** 589–592 (1992).

9. Baird, G. S., Zacharias, D. A. & Tsien, R. Y. Circular permutation and receptor insertion within green fluorescent proteins. *Proc. Natl Acad. Sci. USA* **96,** 11241–11246 (1999).
10. Campbell, R. E. *et al.* A monomeric red fluorescent protein. *Proc. Natl Acad. Sci. USA* **99,** 7877–7882 (2002).
11. Hanson-Smith, V., Kolaczkowski, B. & Thornton, J. W. Robustness of ancestral sequence reconstruction to phylogenetic uncertainty. *Mol. Biol. Evol.* **27,** 1988–1999 (2010).
12. Matsumoto, T., Akashi, H. & Yang, Z. Evaluation of ancestral sequence reconstruction methods to infer nonstationary patterns of nucleotide substitution. *Genetics* **200,** 873–890 (2015).
13. Williams, P. D., Pollock, D. D., Blackburne, B. P. & Goldstein, R. A. Assessing the accuracy of ancestral protein reconstruction methods. *PLoS Comput. Biol.* **2,** e69 (2006).
14. Yang, Z. Among-site rate variation and its impact on phylogenetic analyses. *Trends Ecol. Evol.* **11,** 367–372 (1996).
15. Lartillot, N., Lepage, T. & Blanquart, S. PhyloBayes 3: a Bayesian software package for phylogenetic reconstruction and molecular dating. *Bioinformatics* **25,** 2286–2288 (2009).
16. Groussin, M. *et al.* Toward more accurate ancestral protein genotype-phenotype reconstructions with the use of species tree-aware gene trees. *Mol. Biol. Evol.* **32,** 13–22 (2015).
17. Alieva, N. O. *et al.* Diversity and evolution of coral fluorescent proteins. *PLoS ONE* **3,** e2680 (2008).
18. Ashkenazy, H. *et al.* FastML: a web server for probabilistic reconstruction of ancestral sequences. *Nucleic Acids Res.* **40,** W580–W584 (2012).
19. Yang, Z., Kumar, S. & Nei, M. A new method of inference of ancestral nucleotide and amino acid sequences. *Genetics* **141,** 1641–1650 (1995).
20. Pollock, D. D., Zwickl, D. J., McGuire, J. A. & Hillis, D. M. Increased taxon sampling is advantageous for phylogenetic inference. *Syst. Biol.* **51,** 664–671 (2002).
21. Hillis, D. M. Taxonomic sampling, phylogenetic accuracy, and investigator bias. *Syst. Biol.* **47,** 3–8 (1998).
22. Kim, J. Large-scale phylogenies and measuring the performance of phylogenetic estimators. *Syst. Biol.* **47,** 43–60 (1998).
23. Rosenberg, M. S. & Kumar, S. Incomplete taxon sampling is not a problem for phylogenetic inference. *Proc. Natl Acad. Sci. USA* **98,** 10751–10756 (2001).
24. Duchemin, W., Daubin, V. & Tannier, E. Reconstruction of an ancestral *Yersinia pestis* genome and comparison with an ancient sequence. *BMC Genomics* **16**(Suppl 10): S9 (2015).
25. Yang, K., Heath, L. S. & Setubal, J. C. REGEN: ancestral genome reconstruction for bacteria. *Genes (Basel)* **3,** 423–443 (2012).
26. Yang, N., Hu, F., Zhou, L. & Tang, J. Reconstruction of ancestral gene orders using probabilistic and gene encoding approaches. *PLoS ONE* **9,** e108796 (2014).
27. Chi, P. B. & Liberles, D. A. Selection on protein structure, interaction, and sequence. *Protein Sci.* **25,** 1168–1178 (2016).
28. Pollock, D. D., Thiltgen, G. & Goldstein, R. A. Amino acid coevolution induces an evolutionary Stokes shift. *Proc. Natl Acad. Sci. USA* **109,** E1352–E1359 (2012).
29. Bar-Rogovsky, H. *et al.* Assessing the prediction fidelity of ancestral reconstruction by a library approach. *Protein Eng. Des. Sel.* **28,** 507–518 (2015).
30. Cole, M. F. & Gaucher, E. A. Utilizing natural diversity to evolve protein function: applications towards thermostability. *Curr. Opin. Chem. Biol.* **15,** 399–406 (2011).
31. Cole, M. F. & Gaucher, E. A. Exploiting models of molecular evolution to efficiently direct protein engineering. *J. Mol. Evol.* **72,** 193–203 (2011).
32. Shaner, N. C. *et al.* Improved monomeric red, orange and yellow fluorescent proteins derived from *Discosoma* sp. red fluorescent protein. *Nat. Biotechnol.* **22,** 1567–1572 (2004).
33. Darriba, D., Taboada, G. L., Doallo, R. & Posada, D. ProtTest 3: fast selection of best-fit models of protein evolution. *Bioinformatics* **27,** 1164–1165 (2011).
34. Ai, H. W., Baird, M. A., Shen, Y., Davidson, M. W. & Campbell, R. E. Engineering and characterizing monomeric fluorescent proteins for live-cell imaging applications. *Nat. Protoc.* **9,** 910–928 (2014).
35. Wurth, C., Grabolle, M., Pauli, J., Spieles, M. & Resch-Genger, U. Relative and absolute determination of fluorescence quantum yields of transparent samples. *Nat. Protoc.* **8,** 1535–1550 (2013).

Acknowledgements

Funding was provided by the Georgia Institute of Technology, NASA (NNX12AI10G to E.A.G.), DuPont (Young Professor Award to E.A.G.) and NSF (grant 1145698 to E.A.G) We thank Andreas S. Bommarius for the *mRFP1* gene, Barry G. Hall, Nathan Shaner, Joshua Weitz, and Zihang Yang for scientific discussions, and the following for assistance with building the experimental phylogeny: Kayla Arroyo, Krutika Gaonkar, Kristen Ingram, Penelope Kahn, Mark Leber, Byron Lee, Dione McKenzie, Angeline Pham, Lily Tran, Rebecca Wolf, and Zi-Ming Zhao.

Author contributions

R.N.R. and E.A.G. conceived of the project and wrote the manuscript; R.N.R., C.E.R. and E.A.G. analysed results; R.N.R., C.E.R., K.A.R. and D.K.N performed the experiments.

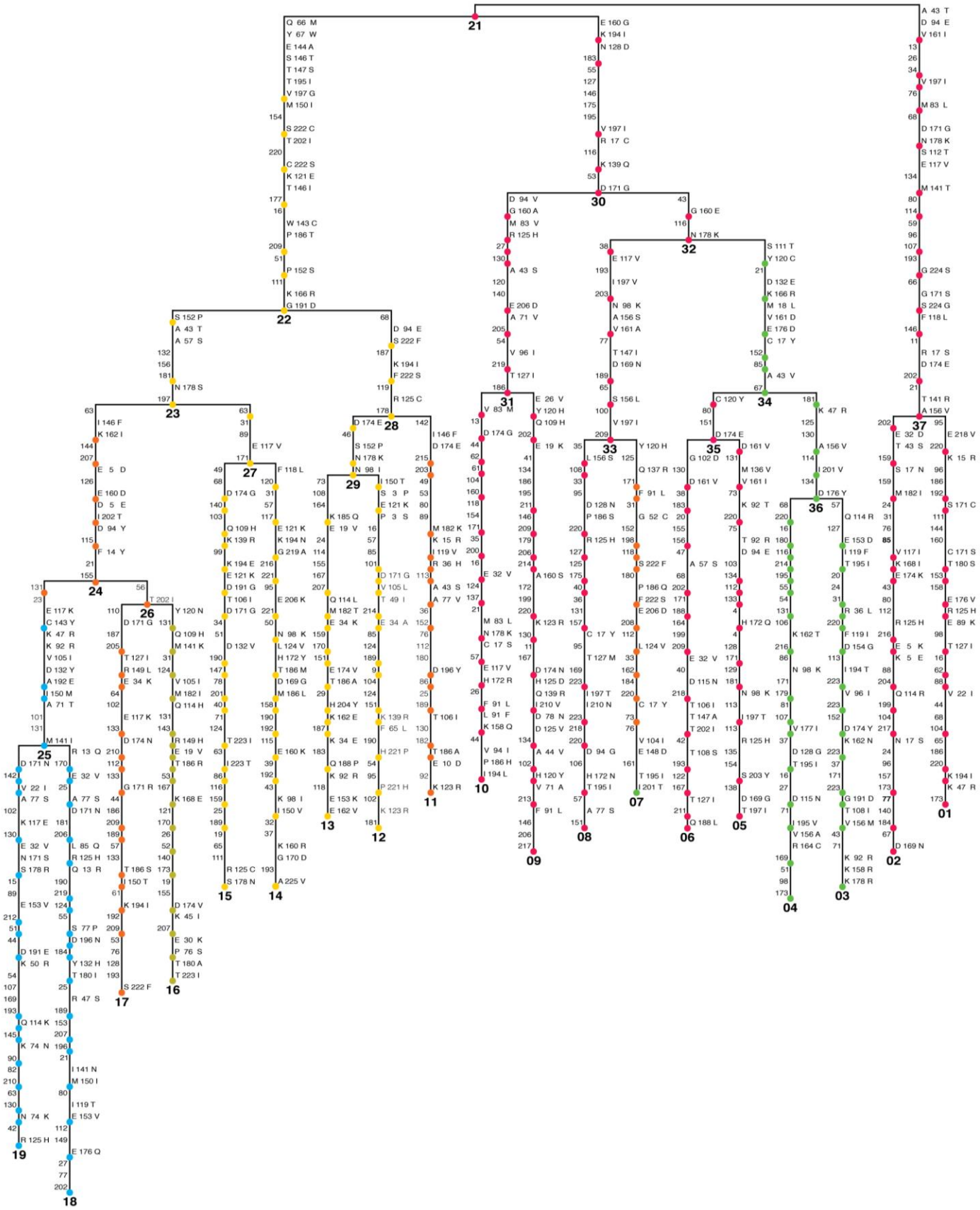
Additional information

Supplementary Information accompanies this paper at http://www.nature.com/naturecommunications

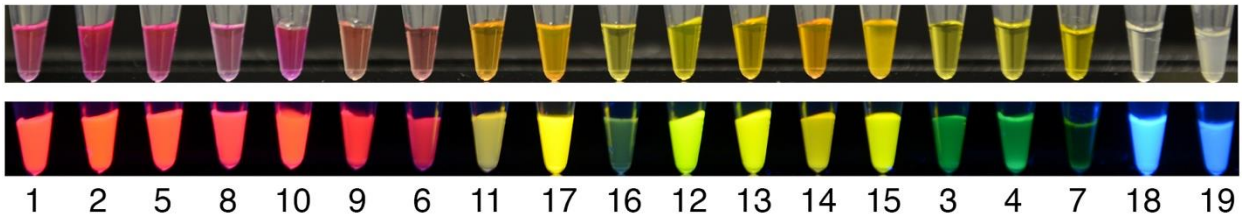
Competing financial interests: The authors declare no competing financial interests.

Reprints and permission information is available online at http://npg.nature.com/reprintsandpermissions/

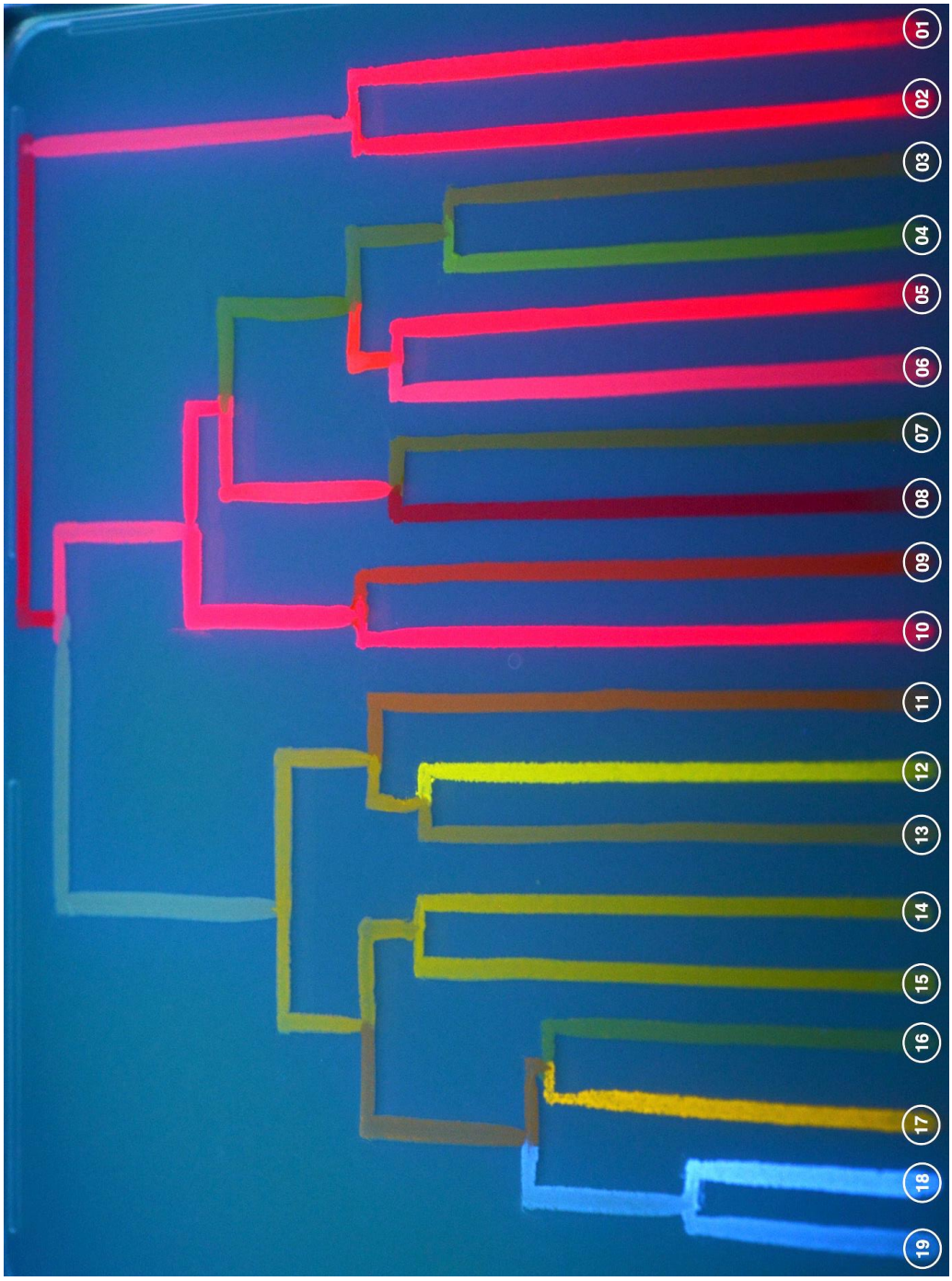
How to cite this article: Randall, R. N. *et al.* An experimental phylogeny to benchmark ancestral sequence reconstruction. *Nat. Commun.* 7:12847 doi: 10.1038/ncomms12847 (2016).



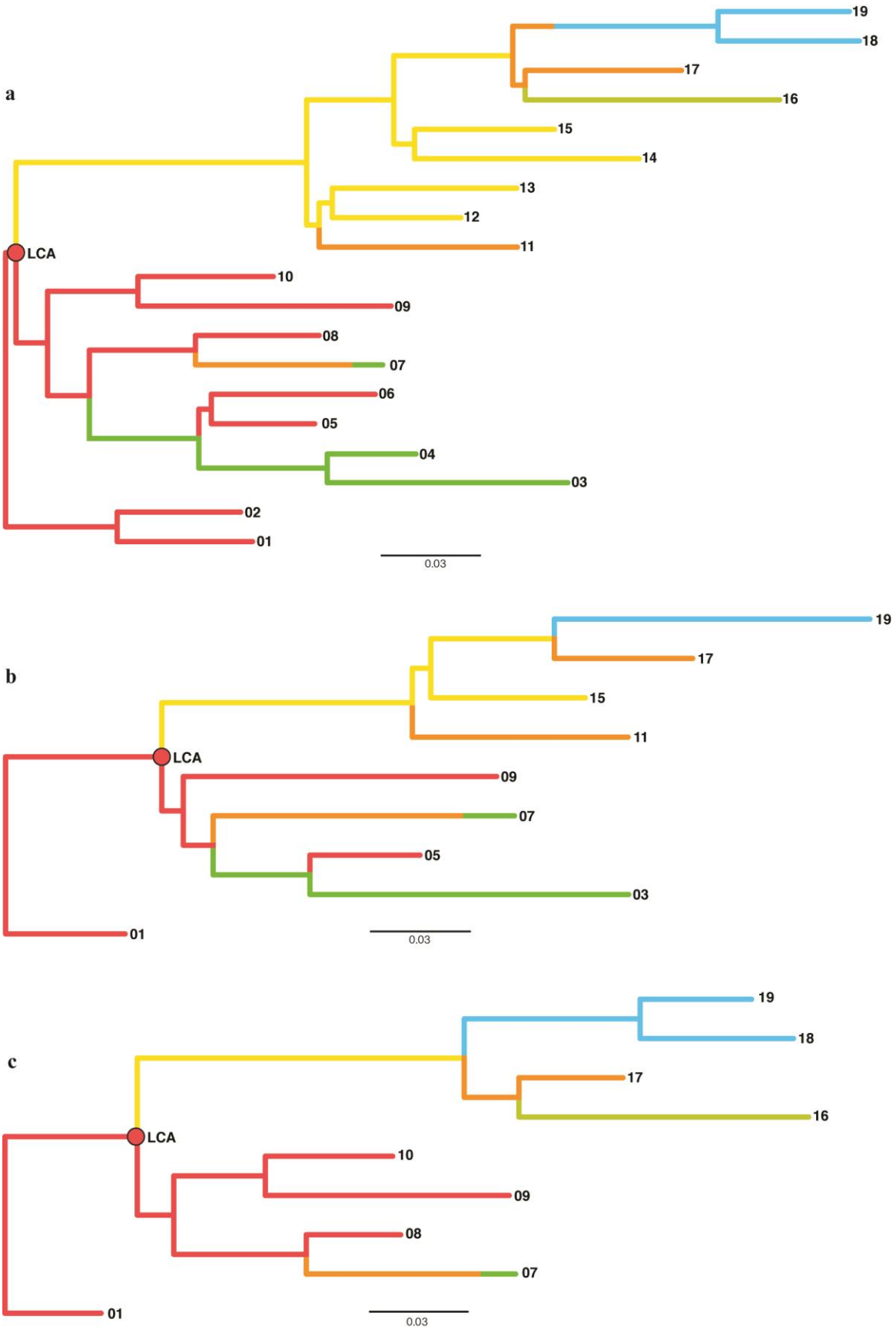
Supplementary Figure 1. Distribution of mutations in the experimental phylogeny. Leaf and node numbers are provided at each tip and bifurcation point on the tree, respectively. Nonsynonymous mutations are listed to the right of each branch and synonymous mutations to the left. Nonsynonymous mutations also list the amino acid replacement. Each filled circle represents one round of random-mutagenesis PCR and the color of the circle represents the color-class phenotype of the FP protein at that location in the tree.



Supplementary Figure 2. Diversity of phenotypes displayed by the leaf ('modern') sequences of the phylogeny. Purified proteins from sequences 1-19 are arranged by color-class phenotype and are shown in visible light (top image) and 365 nm UV light (bottom image).



Supplementary Figure 3. Evolution of color emission in the experimental phylogeny. Cladogram drawn with bacteria growing on an agar plate expressing either a node or leaf fluorescent protein and visualized under 365 nm ultraviolet light. Topology is a representation of the true topology shown in the phylogram (Supplementary Fig. 1).

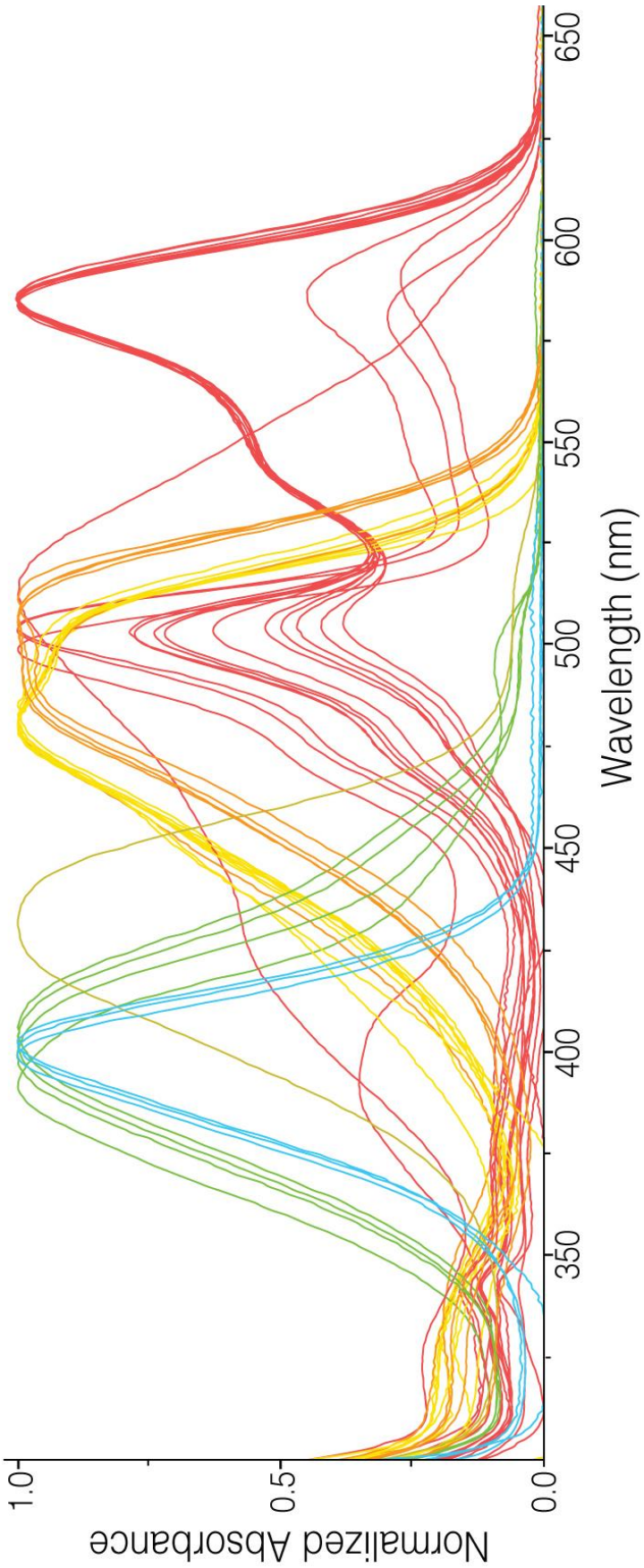

a
19
18
17
16
15
14
13
12
11
LCA
10
09
08
07
06
05
04
03
02
01
0.03
b
19
17
15
11
LCA
09
07
05
03
01
0.03
c
19
18
17
16
LCA
10
09
08
07
01
0.03

d

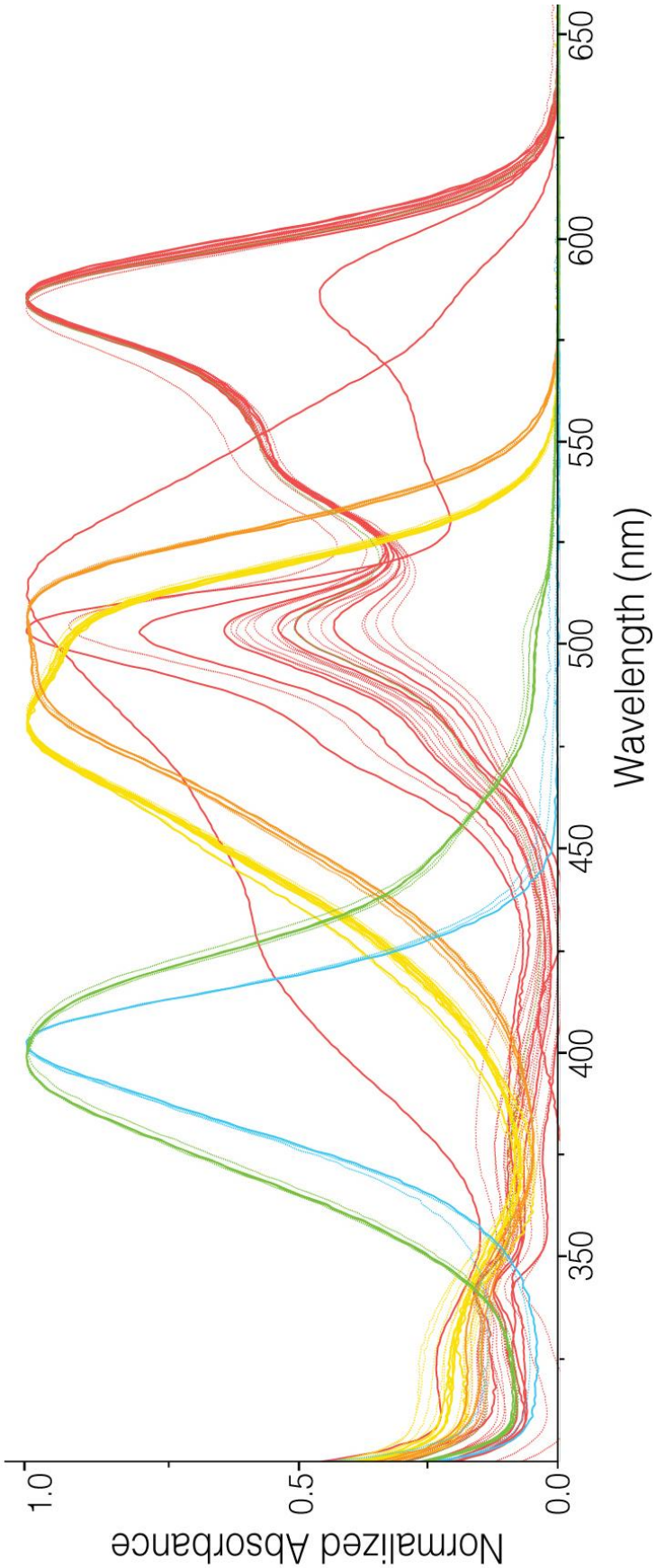
```
TrueLCA          MASSEDVIKEFMRFKVRMEGSVNGHEFEIEGEGEGRPYEGTQAAKLKVTKGGPLPFAWDI
PAML all 19      MASSEDVIKEFMRFKVSMEGSVNGHEFEIEGEGEGRPYEGTQSAKLKVTKGGPLPFAWDI
PAML Sub1        MASSEDVIKEFMRFKVSMEGSVNGHEFEIEGEGEGRPYEGTQTAKLKVTKGGPLPFAWDI
PAML Sub2        MASSEDVIKEFMRFKVSMEGSVNGHEFEIEGEGEGRPYEGTQHAKLKVTKGGPLPFAWDI
                 *************** ************************:*****************

TrueLCA          LSPQFQYGSKAYVKHPADIPDYMKLSFPEGFKWDRVMNFEDGGVVTVTQDSSLQDGEFIY
PAML all 19      LSPQFQYGSKAYVKHPADIPDYMKLSFPEGFKWERVMNFEDGGVVTVTQDSSLQDGVFIY
PAML Sub1        LSPQFQYGSKAYVKHPADIPDYMKLSFPEGFKWERVMNFEDGGVVTVTQDSSLQDGEFIY
PAML Sub2        LSPQFQYGSKAYVKHPADIPDYMKLSFPEGFKWERVMNFEDGGVVTVTQDSSLQDGVFIY
                 **********************************:********************** ***

TrueLCA          KVKLRGTNFPSDGPVMQKKTMGWEASTERMYPEDGALKGEVKMRLKLKDGDHYDAEVNTT
PAML all 19      KVKLHGTNFPSDGPVMQKKTMGWEASTERMYPEDGALKGEVKMRLKLKDGGHYDAEVKTT
PAML Sub1        KVKLHGTNFPSDGPVMQKKTMGWEASTERMYPEDGALKGEVKMRLKLKDGGHYDAEVKTT
PAML Sub2        KVKLHGTNFPSDGPVMQKKTMGWEASTERMYPEDGALKGEVKMRLKLKDGSHYDAEVKTT
                 ****.* ****************************************.**:***:**

TrueLCA          YMAKKPVQLPGAYKTDVKLDITSHNEDYTIVEQYERAEGRHSTGA
PAML all 19      YMAKKPVQLPGAYITDIKLDITSHNEDYTIVEQYERAEGRHSTGA
PAML Sub1        YMAKKPVQLPGAYITDIKLDITSHNEDYTIVEQYERAEGRHSTGA
PAML Sub2        YMAKKPVQLPGAYITDIKLDITSHNEDYTIVEQYERAEGRHSTGA
                 ************* **:****************************
```

Supplementary Figure 4. The effects of taxon sampling on the inference of the last common ancestral sequence. (**a**) Full phylogeny from Fig. 1 in the main article. (**b**) Subsample of sequences across the entire spectrum of the full phylogeny used to infer the last common ancestor (LCA). (**c**) Subsample of sequences composed of clusters from the full phylogeny used to infer the LCA. Scale bars represents amino acid replacements per site per unit evolutionary time. Color corresponds to Fig.1 from the main article. (**d**) Multiple sequence alignment of LCAs from the different analyses; true ancestor, PAML inferred using all of the 19 leaf sequences, PAML using sequences from **b** (Subsample1), PAML using sequences from **c** (Subsample2). Inferred residues that differ from the true ancestral residues are highlighted in magenta. The sites of the highlighted positions are 17, 43, 94, 117, 125, 127, 171, 174, 178, 194 and 197.

a


Normalized Absorbance
0.0
0.5
1.0
350
400
450
500
550
600
650
Wavelength (nm)

b


Normalized Absorbance
0.0
0.5
1.0
350
400
450
500
550
600
650
Wavelength (nm)

Supplementary Figure 5. Absorbance spectra for all 70 proteins characterized in the study. (**a**) Absorbance spectra for true ancestors and leaves of phylogeny. Color of spectral line represents color-class of encoded protein. (**b**) Absorbance spectra for true ancestors (solid lines) and incorrectly inferred ancestors (dotted lines).

Supplementary Table 1. Summary statistics for the experimental phylogeny. (**a**) Total mutations and types of mutations accumulated in the phylogeny. (**b**) A list of specific nucleotide substitution types and occurrences in the phylogeny. Abbreviations ts and tv represent transitions and transversions, repectively. (**c**) Types of nucleotide substitutions resulting in either a synonymous (syn) or nonsynonymous (nonsyn) mutation along each individual branch of the phylogeny. Predicted ratio of nonsynonymous to synonymous mutations does not match the observed ratio because of reversion mutations.

a

Experimental Phylogeny Mutation Summary	
Total variants in phylogeny	349
Nodes	17
Leaves	19
Total Mutations	833
Transitions (ts)	535
Transversions (tv)	298
Total synonymous (syn) mutations	461
Synonymous reversions	14
Observed synonymous mutations	447
Total nonsynonymous (nonsyn) mutations	372
Nonsynonymous reversions	19
Observed nonsynonymous mutations	353

b

Nucleotide Substitution			
Substitution	Type	Occurrence	Percentage (%)
A --> T	tv	74	8.88
A --> C	tv	16	1.92
A --> G	ts	117	14.05
T --> A	tv	72	8.64
T --> C	ts	106	12.73
T --> G	tv	19	2.28
C --> T	ts	167	20.05
C --> G	tv	17	2.04
C --> A	tv	33	3.96
G --> C	tv	21	2.52
G --> T	tv	46	5.52
G --> A	ts	145	17.41

c

Branch number	Syn mutations	Syn reversions	Observed syn mutations	Nonsyn mutations	Nonsyn reversions	Observed nonsyn mutations	Observed nonsyn/syn	Predicted nonsyn/syn
1	21	0	21	12	1	11	0.5714	0.5238
2	24	1	23	13	2	11	0.5417	0.4783
3	16	2	14	17	1	16	1.0625	1.1429
4	25	0	25	9	0	9	0.3600	0.3600
5	19	1	18	13	1	12	0.6842	0.6667
6	21	0	21	11	0	11	0.5238	0.5238
7	16	0	16	14	1	13	0.8750	0.8125
8	21	1	20	12	0	12	0.5714	0.6000
9	23	2	21	16	1	15	0.6957	0.7143
10	18	1	17	14	1	13	0.7778	0.7647
11	19	0	19	13	0	13	0.6842	0.6842
12	17	0	17	13	2	11	0.7647	0.6471
13	15	0	15	14	1	13	0.9333	0.8667
14	19	2	17	17	0	17	0.8947	1.0000
15	23	0	23	13	1	12	0.5652	0.5217
16	16	1	15	16	0	16	1.0000	1.0667
17	22	1	21	11	0	11	0.5000	0.5238
18	21	1	20	17	1	16	0.8095	0.8000
19	19	1	18	14	1	13	0.7368	0.7222
22	7	0	7	20	1	19	2.8571	2.7143
23	4	0	4	4	0	4	1.0000	1.0000
24	7	0	7	8	1	7	1.1429	1.0000
25	4	0	4	10	0	10	2.5000	2.5000
26	1	0	1	1	0	1	1.0000	1.0000
27	4	0	4	1	0	1	0.2500	0.2500
28	4	0	4	5	1	4	1.2500	1.0000
29	1	0	1	4	0	4	4.0000	4.0000
30	8	0	8	7	0	7	0.8750	0.8750
31	8	0	8	9	0	9	1.1250	1.1250
32	2	0	2	2	0	2	1.0000	1.0000
33	8	0	8	9	1	8	1.1250	1.0000
34	4	0	4	9	0	9	2.2500	2.2500
35	2	0	2	2	0	2	1.0000	1.0000
36	5	0	5	4	0	4	0.8000	0.8000
37	17	0	17	18	1	17	1.0588	1.0000
						Average	1.0510	1.0267

Supplementary Table 2. Percentage of sites correctly inferred for the five tested ASR procedures. Procedures follow the color-code scheme (Fig. 2 in main article).

Method	Correctly Inferred Sites	Incorrectly Inferred Sites	Total Sites
PAML_Γ	98.14%	71	3825
FASTML_Γ	98.17%	70	3825
PAML	98.12%	72	3825
PHYLO_Γ	97.88%	81	3825
MP	98.07%	74	3825

Supplementary Table 3. (**a**) Phenotypes of true ancestors and incorrectly inferred ancestors. Extinction coefficient (ε), quantum yield (Φ), brightness (ε x Φ), excitation maxima (λ_{ex}), and emission maxima (λ_{em}). The percent error equals zero when the inferred sequence is identical to the ancestral sequence. (**b**) Phenotypes for all 19 leaf proteins. (**c**) Summary of phenotypic error for each ASR procedure. Note, nodes 34 and 35 are not included because they outweigh all other error (but their errors are the same for the five procedures so the overall effect is negligible. Procedures follow the color-code scheme (Supplementary Table 2a, c, Fig. 2).

a

Node	Σ ($M^{-1}cm^{-1}$)	Σ Error	√	√ Error	λ_{ex} (nm)	λ_{ex} Error	λ_{em} (nm)	λ_{em} Error	Brightness	Brightness Error
21	19,003	-	0.1213	-	586	-	608	-	2,306	-
PAML_Γ	36,317	91.11%	0.2567	111.53%	585	0.17%	604	0.66%	9,322	304.25%
FASTML_Γ	36,317	91.11%	0.2567	111.53%	585	0.17%	604	0.66%	9,322	304.25%
PAML	36,317	91.11%	0.2567	111.53%	585	0.17%	604	0.66%	9,322	304.25%
PHYLO_Γ	36,317	91.11%	0.2567	111.53%	585	0.17%	604	0.66%	9,322	304.25%
MP	46,312	143.71%	0.2532	108.62%	585	0.17%	606	0.33%	11,724	408.43%
22	7,307	-	0.0710	-	486	-	549	-	519	-
PAML_Γ	18,524	153.51%	0.0659	7.22%	486	0.00%	546	0.55%	1,221	135.21%
FASTML_Γ	18,524	153.51%	0.0659	7.22%	486	0.00%	546	0.55%	1,221	135.21%
PAML	19,352	164.84%	0.0768	8.09%	486	0.00%	546	0.55%	1,486	186.26%
PHYLO_Γ	19,352	164.84%	0.0768	8.09%	486	0.00%	546	0.55%	1,486	186.26%
MP	18,276	150.12%	0.0656	7.60%	486	0.00%	546	0.55%	1,199	131.11%
23	16,329	-	0.0803	-	486	-	546	-	1,312	-
PAML_Γ	16,443	0.70%	0.0770	4.18%	486	0.00%	546	0.00%	1,266	3.51%
FASTML_Γ	16,443	0.70%	0.0770	4.18%	486	0.00%	546	0.00%	1,266	3.51%
PAML	20,125	23.25%	0.0809	0.74%	486	0.00%	546	0.00%	1,628	24.16%
PHYLO_Γ	20,125	23.25%	0.0809	0.74%	486	0.00%	546	0.00%	1,628	24.16%
MP	18,353	12.40%	0.0773	3.71%	486	0.00%	546	0.00%	1,419	8.23%
24	24,473	-	0.0713	-	512	-	560	-	1,745	-
PAML_Γ	26,404	7.89%	0.0831	16.49%	511	0.20%	560	0.00%	2,193	25.69%
FASTML_Γ	26,404	7.89%	0.0831	16.49%	511	0.20%	560	0.00%	2,193	25.69%
PAML	26,404	7.89%	0.0831	16.49%	511	0.20%	560	0.00%	2,193	25.69%
PHYLO_Γ	25,695	4.99%	0.0805	12.87%	512	0.00%	561	0.18%	2,068	18.50%
MP	25,486	4.14%	0.0769	7.93%	512	0.00%	560	0.00%	1,961	12.40%
25	17,857	-	0.0868	-	373	-	452	-	1,550	-
PAML_Γ	11,018	38.30%	0.1036	19.39%	399	6.97%	451	0.22%	1,142	26.34%
FASTML_Γ	11,018	38.30%	0.1036	19.39%	399	6.97%	451	0.22%	1,142	26.34%
PAML	24,857	39.20%	0.0989	13.90%	400	7.24%	451	0.22%	2,458	58.55%
PHYLO_Γ	24,857	39.20%	0.0989	13.90%	400	7.24%	451	0.22%	2,458	58.55%
MP	11,018	38.30%	0.1036	19.39%	399	6.97%	451	0.22%	1,142	26.34%
26	25,592	-	0.0720	-	512	-	560	-	1,842	-
PAML_Γ	25,592	0.00%	0.0720	0.00%	512	0.00%	560	0.00%	1,842	0.00%
FASTML_Γ	25,592	0.00%	0.0720	0.00%	512	0.00%	560	0.00%	1,842	0.00%
PAML	22,798	10.92%	0.0756	5.08%	511	0.20%	560	0.00%	1,725	6.39%
PHYLO_Γ	26,469	3.43%	0.0708	1.64%	512	0.00%	560	0.00%	1,874	1.74%
MP	25,486	0.41%	0.0769	6.89%	512	0.00%	560	0.00%	1,961	6.45%
27	18,419	-	0.0758	-	486	-	546	-	1,396	-
PAML_Γ	20,125	9.26%	0.0809	6.78%	486	0.00%	546	0.00%	1,628	16.67%
FASTML_Γ	20,125	9.26%	0.0809	6.78%	486	0.00%	546	0.00%	1,628	16.67%
PAML	20,125	9.26%	0.0809	6.78%	486	0.00%	546	0.00%	1,628	16.67%
PHYLO_Γ	20,125	9.26%	0.0809	6.78%	486	0.00%	546	0.00%	1,628	16.67%
MP	16,939	8.04%	0.0744	1.75%	486	0.00%	546	0.00%	1,261	9.64%
28	17,288	-	0.0769	-	486	-	545	-	1,329	-
PAML_Γ	18,524	7.15%	0.0659	14.28%	486	0.00%	546	0.18%	1,221	8.15%
FASTML_Γ	18,524	7.15%	0.0659	14.28%	486	0.00%	546	0.18%	1,221	8.15%
PAML	18,524	7.15%	0.0659	14.28%	486	0.00%	546	0.18%	1,221	8.15%
PHYLO_Γ	15,953	7.72%	0.0753	2.01%	486	0.00%	546	0.18%	1,202	9.58%

MP	18,276	5.71%	0.0656	14.63%	486	0.00%	546	0.18%	1,199	9.75%
29	18,034	-	0.0803	-	487	-	547	-	1,448	-
PAML_Γ	17,102	5.17%	0.0775	3.54%	487	0.00%	547	0.00%	1,325	8.53%
FASTML_Γ	17,102	5.17%	0.0775	3.54%	487	0.00%	547	0.00%	1,325	8.53%
PAML	17,102	5.17%	0.0775	3.54%	487	0.00%	547	0.00%	1,325	8.53%
PHYLO_Γ	19,237	6.67%	0.0815	1.43%	486	0.21%	547	0.00%	1,567	8.20%
MP	18,034	0.00%	0.0803	0.00%	487	0.00%	547	0.00%	1,448	0.00%
30	26,360	-	0.2398	-	586	-	606	-	6,322	-
PAML_Γ	32,151	21.97%	0.2486	3.67%	586	0.00%	604	0.33%	7,993	26.44%
FASTML_Γ	32,151	21.97%	0.2486	3.67%	586	0.00%	604	0.33%	7,993	26.44%
PAML	32,151	21.97%	0.2486	3.67%	586	0.00%	604	0.33%	7,993	26.44%
PHYLO_Γ	37,090	40.71%	0.2489	3.77%	585	0.17%	605	0.17%	9,231	46.01%
MP	39,835	51.12%	0.2547	6.19%	586	0.00%	605	0.17%	10,145	60.48%
31	36,941	-	0.2528	-	584	-	602	-	9,339	-
PAML_Γ	33,535	9.22%	0.2514	0.54%	584	0.00%	604	0.33%	8,431	9.71%
FASTML_Γ	33,535	9.22%	0.2514	0.54%	584	0.00%	604	0.33%	8,431	9.71%
PAML	45,841	24.09%	0.2437	3.60%	585	0.17%	605	0.50%	11,171	19.62%
PHYLO_Γ	45,841	24.09%	0.2437	3.60%	585	0.17%	605	0.50%	11,171	19.62%
MP	49,196	33.17%	0.2527	0.03%	585	0.17%	605	0.50%	12,433	33.14%
32	38,038	-	0.2527	-	586	-	608	-	9,614	-
PAML_Γ	31,981	15.92%	0.2483	1.76%	585	0.17%	604	0.66%	7,941	17.40%
FASTML_Γ	31,981	15.92%	0.2483	1.76%	585	0.17%	604	0.66%	7,941	17.40%
PAML	31,981	15.92%	0.2483	1.76%	585	0.17%	604	0.66%	7,941	17.40%
PHYLO_Γ	28,842	24.18%	0.2616	3.52%	585	0.17%	605	0.49%	7,546	21.51%
MP	31,981	15.92%	0.2483	1.76%	585	0.17%	604	0.66%	7,941	17.40%
33	30,069	-	0.2572	-	584	-	604	-	7,733	-
PAML_Γ	30,873	2.67%	0.2178	15.31%	583	0.17%	604	0.00%	6,724	13.05%
FASTML_Γ	30,873	2.67%	0.2178	15.31%	583	0.17%	604	0.00%	6,724	13.05%
PAML	30,873	2.67%	0.2178	15.31%	583	0.17%	604	0.00%	6,724	13.05%
PHYLO_Γ	30,873	2.67%	0.2178	15.31%	583	0.17%	604	0.00%	6,724	13.05%
MP	30,873	2.67%	0.2178	15.31%	583	0.17%	604	0.00%	6,724	13.05%
34	26,491	-	0.0019	-	402	-	512	-	51	-
PAML_Γ	30,213	14.05%	0.2326	11927.88%	584	45.27%	604	17.97%	7,028	13617.80%
FASTML_Γ	30,213	14.05%	0.2326	11927.88%	584	45.27%	604	17.97%	7,028	13617.80%
PAML	30,213	14.05%	0.2326	11927.88%	584	45.27%	604	17.97%	7,028	13617.80%
PHYLO_Γ	30,213	14.05%	0.2326	11927.88%	584	45.27%	604	17.97%	7,028	13617.80%
MP	30,213	14.05%	0.2326	11927.88%	584	45.27%	604	17.97%	7,028	13617.80%
35	8,629	-	0.0069	-	585	-	605	-	60	-
PAML_Γ	30,213	250.13%	0.2326	3270.90%	584	0.17%	604	0.17%	7,028	11702.63%
FASTML_Γ	30,213	250.13%	0.2326	3270.90%	584	0.17%	604	0.17%	7,028	11702.63%
PAML	30,213	250.13%	0.2326	3270.90%	584	0.17%	604	0.17%	7,028	11702.63%
PHYLO_Γ	31,380	263.66%	0.2481	3496.06%	583	0.34%	603	0.33%	7,787	12977.33%
MP	30,213	250.13%	0.2326	3270.90%	584	0.17%	604	0.17%	7,028	11702.63%
36	24,031	-	0.0010	-	406	-	511	-	23	-
PAML_Γ	25,263	5.13%	0.0010	1.54%	400	1.48%	511	0.00%	25	6.75%
FASTML_Γ	27,819	15.76%	0.0011	12.79%	400	1.48%	511	0.00%	30	30.57%
PAML	27,819	15.76%	0.0011	12.79%	400	1.48%	511	0.00%	30	30.57%
PHYLO_Γ	27,819	15.76%	0.0011	12.79%	400	1.48%	511	0.00%	30	30.57%
MP	27,266	13.46%	0.0021	119.23%	400	1.48%	511	0.00%	57	148.74%
37	39,333	-	0.2905	-	584	-	604	-	11,425	-
PAML_Γ	49,835	26.70%	0.2904	0.04%	584	0.00%	603	0.17%	14,470	26.65%
FASTML_Γ	49,835	26.70%	0.2904	0.04%	584	0.00%	603	0.17%	14,470	26.65%
PAML	49,835	26.70%	0.2904	0.04%	584	0.00%	603	0.17%	14,470	26.65%
PHYLO_Γ	49,835	26.70%	0.2904	0.04%	584	0.00%	603	0.17%	14,470	26.65%
MP	49,835	26.70%	0.2904	0.04%	584	0.00%	603	0.17%	14,470	26.65%

b

Leaf	$\varepsilon(M^{-1}cm^{-1})$	Φ	λ_{ex} (nm)	λ_{em} (nm)	Brightness
1	43,663	0.2819	585	604	12,309
2	50,293	0.2837	584	602	14,270
3	26,750	0.0036	398	511	96
4	26,759	0.0052	425	513	140
5	61,727	0.2717	584	605	16,773
6	38,793	0.2541	584	605	9,858
7	30,084	0.0008	497	511	24
8	19,474	0.0747	580	604	1,455
9	8,572	0.2321	589	609	1,990
10	46,478	0.2412	584	603	11,208
11	27,193	0.0544	509	557	1,480
12	17,339	0.1335	486	532	2,314
13	11,352	0.0716	486	548	813
14	17,366	0.0951	486	543	1,651
15	13,477	0.0939	486	533	1,265
16	23,978	0.0061	510	565	145
17	25,406	0.1648	512	556	4,186
18	29,356	0.1348	398	451	3,956
19	10,572	0.1038	399	450	1,097

c

Method	**ε Average Error**	**Φ Average Error**	**Brightness Average Error**
PAML_Γ	26.31%	13.75%	41.89%
FASTML_Γ	27.02%	14.50%	43.48%
PAML	31.06%	14.51%	51.49%
PHYLO_Γ	32.31%	13.20%	52.36%
MP	33.73%	20.87%	60.79%

Nodes 34 and 35 not included due to being outliers

Supplementary Note 1

Amino acid sequences for the 37 proteins at the leaves and nodes in the experimental phylogeny.

>01
MASSEDVIKEFMRFRVSMEGSINGHEFEIEGEGEGRPYEGTQTAKLRVTKGGPLPFAWDILSPQF
QYGSKAYVKHPADIPDYLKLSFPKGFKWERVMNFEDGGVVTVTQDSTLQDGVLIYKVKLHGINFP
SDGPVMQKKTRGWEASTERMYPEDGVLKGEIKMRLKLKDGSHYEAVVKTSYMAKKPVQLPGAYIT
DIKLDITSHNEDYTIVEQYERAVGRHSTGA

>02
MASSEDVIKEFMRFKVSMEGSVNGHEFEIEGDGEGRPYEGTQSAKLKVTKGGPLPFAWDILSPQF
QYGSKAYVKHPADIPDYLKLSFPEGFKWERVMNFEDGGVVTVTQDSTLRDGILIYKVKLHGTNFP
SDGPVMQKKTRGWEASTERMYPEDGVLKGEIKMRLKLINGSHYKAEVKTTYIAKKPVQLPGAYKT
DIKLDITSHNEDYTIVEQYERAEGRHSTGA

>03
MASSEDVIKEFMRFKVYLEGSVNGHEFEIEGEGEGLPYEGTQVAKLRVTKGGPLPFAWDILSPQF
QYGSKAYVKHPADIPDYMKLSFPEGFRWDRIMNFEDGGVVTVIQDTSLRDGEFICKVKLRGTDFP
SEGPVMQKQTMGWEASTERMYPDGGMLRGEDNMRLRLKDGGHYYAYVRTTYMAKKPVQLPDAYTI
DIKLDVTSHNEDYTIVEQYERAEGRHSTGA

>04
MASSEDVIKEFMRFKVYLEGSVNGHEFEIEGEGEGRPYEGTQVAKLRVTKGGPLPFAWDILSPQF
QYGSKAYVKHPADIPDYMKLSFPEGFKWDRVMKFEDGGVVTVTQDTSLQNGEFICKVKLRGTGFP
SEGPVMQKQTMGWEASTERMYPEDGALKGEDTMCLRLKDGGHYDAYIKTTYMAKKPVQLPGAYIV
DIKLDVTSHNEDYTIVEQYERAEGRHSTGA

>05
MASSEDVIKEFMRFKVYLEGSVNGHEFEIEGEGEGRPYEGTQVAKLKVTKGGPLPFAWDILSPQF
QYGSKAYVKHPADIPDYMKLSFPEGFRWERVMKFEDGGVVTVTQDTSLQDGEFIYKVKLHGTDFP
SEGPVVQKQTMGWEASTERMYPEDGALKGEIKMRLRLKGGGQYEADVKTTYMAKKPVQLPGAYIT
DIKLDITYHNEDYTIVEQYERAEGRHSTGA

>06
MASSEDVIKEFMRFKVYLEGSVNGHEFEIEGVGEGRPYEGTQVAKLKVTKGGPLPFSWDILSPQF
QYGSKAYVKHPADIPDYMKLSFPEGFKWDRVMNFEDDGVVIVSQDTSLQNGEFIYKVKLRGIDFP
SEGPVMQKQTMGWEASAERMYPEDGALKGEVKMRLRLKDGGHYEADVKTTYMAKKPVLLPGAYIT
DIKLDIISHNEDYTIVEQYERAEGRHSTGA

>07
MASSEDVIKEFMRFKVYMEGSVNGHEFEIEGEGEGRPYEGTQAAKLKVTKGCPLPFAWDILSPQF
QYGSKAYVKHPADIPDYMKLSFPEGLKWDRVMKFEDGGIVTVTQDSSLQDGVFIHKVKVRGTDFP
SDGPVMRKQTMGWEASIDRMYPEDGLLKGEAKMRLKLKNGGHYDAEVKTTYMAKKQVQLPGAYII
DIKLDTTSHNDDYTIVEQYERAEGRHSTGA

>08
MASSEDVIKEFMRFKVYMEGSVNGHEFEIEGEGEGRPYEGTQAAKLKVTKGGPLPFAWDILSPQF
QYGSKAYVKHPSDIPDYMKLSFPEGFKWGRVMKFEDGGVVTVTQDSSLQDGVFIYKVKLHGMNFP
SDGPVMQKQTMGWEASIERMYPEDGSLKGEAKMRLKLKNGGNYDAEVKTTYMAKKSVQLPGAYII
DTKLDITSHNEDYTNVEQYERAEGRHSTGA

>09
MASSEDVIKEFMRFKVCMKGSVNGHVFEIEGEGEGRPYEGTQSVKLKVTKGGPLPFAWDILSPQF
QYGSKAYVKHPANIPDYVKLSFPEGLKWVRIMNFEDGGVVTVTHDSSLQDGEFIYKVRLVGIDFP
SDGPVMQKRTMGWEASTERMYPEDGALKGSVKMRLKLKDGGHYNAEVNTTYMAKKPVQLPGAYIT
DIKLDITSHNDDYTVVEQYERAEGRHSTGA

>10
MASSEDVIKEFMRFKVSMEGSVNGHEFEIEGVGEGRPYEGTQSAKLKVTKGGPLPFAWDILSPQF
QYGSKVYVKHPADIPDYLKLSFPEGFKWIRIMNFEDGGVVTVTQDSSLQDGVFIYKVKLHGIDFP
SDGPVMQKQTMGWEASTERMYPEDGALQGAVKMRLKLKDGGRYGAEVKTTYMAKKHVQLPGAYLT
DIKLDITSHNDDYTIVEQYERAEGRHSTGA

>11
MASSEDVIKDFMRFRVRMEGSVNGHEFEIEGEGEGHPYEGTQSAKLKVTKGGPLPFAWDILSPQF
MWGSKAYVKHPVDIPDYMKLSFPEGFKWERVMNFEDGGVVIVTQDSSLQDGEFVYEVRLCGTNFP
SDGPVMQKKTMGCAAFSERIYSEDGALKGEVKMRLRLKDGDHYEAEVNTTYKAKKAVQLPDAYII
YGKLDIISHNEDYTIVEQYERAEGRHSTGA

>12
MASSEDVIKEFMRFKVRMEGSVNGHEFEIEGEGAGRPYEGTQAAKLKVIKGGPLPFAWDILSPQL
MWGSKAYVKHPADIPDYMKLSFPEGFKWERVMIFEDGGVLTVTQDSSLQDGEFIYKVRLCGTNFP
SDGPVMQKRTMGCAAISERTYPEDGALKGEVKMRLRLKDGGHYEAEVKTTYMAKKTVQLPDAYII
DGKLDIISHNEDYTIVEQYERAEGRHSTGA

>13
MASSEDVIKEFMRFKVRMVGSVNGHEFEIEGEGEGRPYEGTQAAKLKVTKGGPLPFAWDILSPQF
MWGSKAYVKHPADIPDYMKLSFPEGFRWERVMIFEDGGVVTVTQDSSLLDGEFIYEVKLCGTNFP
SDGPVMQKKTMGCAAISERIYPKDGALKGEVVMRLRLKDGDHYVAEVKTTYTAKQAVPLPDAYII
DGKLDIISYNEDYTIVEQYERAEGRHSTGA

>14
MASSEDVIKEFMRFKVRMEGSVNGHEFEIEGEGEGRPYEGTQTAKLKVTKGGPLPFSWDILSPQF
MWGSKAYVKHPADIPDYMKLSFPEGFKWDRVMIFEDGGVVTVTQDSSLQDGVLIYKVKVRGTNFP
SDGPVMQKKTMGCAAISERVYPEDGALKGRVKMRLRLKGDDYYDAEVSTTYMAKKLVQLPDAYNI
DGKLDIISHNKDYTIVEQYERAEARHSTGV

>15
MASSEDVIKEFMRFKVRMEGSVNGHEFEIEGEGEGRPYEGTQTAKLKVTKGGPLPFSWDILSPQF
MWGSKAYVKHPADIPDYMKLSFPEGFKWDRVMNFEDGGVVIVTHDSSLQDGVFIYKVKLCGTNFP
SVGPVMQKRTMGCAAISERIYPEDGALKGEVKMRLRLKDGGHYGAEVNTTYMAKKTVQLPGAYEI
DGKLDIISHNEDYTIVEQYERAEGRHSTGA

>16
MASSEDVIKEFMRYKVRMVGSVNGHEFEIKGEGEGRPYEGTQTAILKVTKGGPLPFSWDILSPQF
MWGSKAYVKHSADIPDYMKLSFPEGFKWYRVMNFEDGGVITVTHDSSLHDGEFINEVKLRGTNFP
SDGPVMQKKTKGCAAFSEHIYPEDGALKGDVIMRLRLEDGDHYVAEVSTAYIAKKRVQLPDAYKI
DGKLDIISHNEDYTIVEQYERAEGRHSIGA

>17
MASSEDVIKEFMRYKVRMEGSVNGHEFEIEGEGKGRPYEGTQTAKLKVTKGGPLPFSWDILSPQF
MWGSKAYVKHPADIPDYMKLSFPEGFKWYRVMNFEDGGVVTVTQDSSLQDGKFIYEVKLRGINFP
SDGPVMQKKTMGCAAFSELTYPEDGALKGDVIMRLRLKDGRHYNAEVSTTYMAKKSVQLPDAYII
DGKLDIISHNEDYTIVEQYERAEGRHFTGA

>18
MASSEDVIKEFMRYKVRMEGSVNGHEFEIEGVGEGRPYEGTQTAKLSVTKGGPLPFSWDILSPQF
MWGSKTYVKHPPDIPDYMKQSFPEGFRWYRVMNFEDGGVITVTQDSSLQDGKFTYEVKLHGTNFP
SHGPVMQKKTNGYAAFSERIYPVDGALKGDVIMRLRLKDGNHYDAQVSTIYMAKKTVQLPDEYKI
NGKLDITSHNEDYTIVEQYERAEGRHSTGA

>19
MASSEDVIKEFMRYKVRMEGSINGHEFEIEGVGEGRPYEGTQTAKLRVTRGGPLPFSWDILSPQF
MWGSKTYVKHPSDIPDYMKLSFPEGFRWYRVMNFEDGGVITVTQDSSLKDGEFIYEVKLHGTNFP
SYGPVMQKKTIGYAAFSERMYPVDGALKGDVIMRLRLKDGSHYDAEVRTTYMAKKTVQLPEEYKI
DGKLDITSHNEDYTIVEQYERAEGRHSTGA

>21
MASSEDVIKEFMRFKVRMEGSVNGHEFEIEGEGEGRPYEGTQAAKLKVTKGGPLPFAWDILSPQF
QYGSKAYVKHPADIPDYMKLSFPEGFKWDRVMNFEDGGVVTVTQDSSLQDGEFIYKVKLRGTNFP
SDGPVMQKKTMGWEASTERMYPEDGALKGEVKMRLKLKDGDHYDAEVNTTYMAKKPVQLPGAYKT
DVKLDITSHNEDYTIVEQYERAEGRHSTGA

>22
MASSEDVIKEFMRFKVRMEGSVNGHEFEIEGEGEGRPYEGTQAAKLKVTKGGPLPFAWDILSPQF
MWGSKAYVKHPADIPDYMKLSFPEGFKWDRVMNFEDGGVVTVTQDSSLQDGEFIYEVKLRGTNFP
SDGPVMQKKTMGCAAISERIYSEDGALKGEVKMRLRLKDGDHYDAEVNTTYMAKKTVQLPDAYKI
DGKLDIISHNEDYTIVEQYERAEGRHSTGA

>23
MASSEDVIKEFMRFKVRMEGSVNGHEFEIEGEGEGRPYEGTQTAKLKVTKGGPLPFSWDILSPQF
MWGSKAYVKHPADIPDYMKLSFPEGFKWDRVMNFEDGGVVTVTQDSSLQDGEFIYEVKLRGTNFP
SDGPVMQKKTMGCAAISERIYPEDGALKGEVKMRLRLKDGDHYDAEVSTTYMAKKTVQLPDAYKI
DGKLDIISHNEDYTIVEQYERAEGRHSTGA

>24
MASSEDVIKEFMRYKVRMEGSVNGHEFEIEGEGEGRPYEGTQTAKLKVTKGGPLPFSWDILSPQF
MWGSKAYVKHPADIPDYMKLSFPEGFKWYRVMNFEDGGVVTVTQDSSLQDGEFIYEVKLRGTNFP
SDGPVMQKKTMGCAAFSERIYPEDGALKGDVIMRLRLKDGDHYDAEVSTTYMAKKTVQLPDAYKI
DGKLDITSHNEDYTIVEQYERAEGRHSTGA

>25
MASSEDVIKEFMRYKVRMEGSVNGHEFEIEGEGEGRPYEGTQTAKLRVTKGGPLPFSWDILSPQF
MWGSKTYVKHPADIPDYMKLSFPEGFRWYRVMNFEDGGVITVTQDSSLQDGKFIYEVKLRGTNFP
SYGPVMQKKTIGYAAFSERMYPEDGALKGDVIMRLRLKDGDHYDAEVSTTYMAKKTVQLPDEYKI
DGKLDITSHNEDYTIVEQYERAEGRHSTGA

>26
MASSEDVIKEFMRYKVRMEGSVNGHEFEIEGEGEGRPYEGTQTAKLKVTKGGPLPFSWDILSPQF
MWGSKAYVKHPADIPDYMKLSFPEGFKWYRVMNFEDGGVVTVTQDSSLQDGEFIYEVKLRGTNFP
SDGPVMQKKTMGCAAFSERIYPEDGALKGDVIMRLRLKDGDHYDAEVSTTYMAKKTVQLPDAYKI
DGKLDIISHNEDYTIVEQYERAEGRHSTGA

>27
MASSEDVIKEFMRFKVRMEGSVNGHEFEIEGEGEGRPYEGTQTAKLKVTKGGPLPFSWDILSPQF
MWGSKAYVKHPADIPDYMKLSFPEGFKWDRVMNFEDGGVVTVTQDSSLQDGVFIYEVKLRGTNFP
SDGPVMQKKTMGCAAISERIYPEDGALKGEVKMRLRLKDGDHYDAEVSTTYMAKKTVQLPDAYKI
DGKLDIISHNEDYTIVEQYERAEGRHSTGA

>28
MASSEDVIKEFMRFKVRMEGSVNGHEFEIEGEGEGRPYEGTQAAKLKVTKGGPLPFAWDILSPQF
MWGSKAYVKHPADIPDYMKLSFPEGFKWERVMNFEDGGVVTVTQDSSLQDGEFIYEVKLCGTNFP
SDGPVMQKKTMGCAAISERIYSEDGALKGEVKMRLRLKDGDHYDAEVNTTYMAKKTVQLPDAYII
DGKLDIISHNEDYTIVEQYERAEGRHSTGA

>29
MASSEDVIKEFMRFKVRMEGSVNGHEFEIEGEGEGRPYEGTQAAKLKVTKGGPLPFAWDILSPQF
MWGSKAYVKHPADIPDYMKLSFPEGFKWERVMIFEDGGVVTVTQDSSLQDGEFIYEVKLCGTNFP
SDGPVMQKKTMGCAAISERIYPEDGALKGEVKMRLRLKDGDHYEAEVKTTYMAKKTVQLPDAYII
DGKLDIISHNEDYTIVEQYERAEGRHSTGA

>30
MASSEDVIKEFMRFKVCMEGSVNGHEFEIEGEGEGRPYEGTQAAKLKVTKGGPLPFAWDILSPQF
QYGSKAYVKHPADIPDYMKLSFPEGFKWDRVMNFEDGGVVTVTQDSSLQDGEFIYKVKLRGTDFP
SDGPVMQKQTMGWEASTERMYPEDGALKGGVKMRLKLKDGGHYDAEVNTTYMAKKPVQLPGAYIT
DIKLDITSHNEDYTIVEQYERAEGRHSTGA

>31
MASSEDVIKEFMRFKVCMEGSVNGHEFEIEGEGEGRPYEGTQSAKLKVTKGGPLPFAWDILSPQF
QYGSKVYVKHPADIPDYVKLSFPEGFKWVRIMNFEDGGVVTVTQDSSLQDGEFIYKVKLHGIDFP
SDGPVMQKQTMGWEASTERMYPEDGALKGAVKMRLKLKDGGHYDAEVNTTYMAKKPVQLPGAYIT
DIKLDITSHNDDYTIVEQYERAEGRHSTGA

>32
MASSEDVIKEFMRFKVCMEGSVNGHEFEIEGEGEGRPYEGTQAAKLKVTKGGPLPFAWDILSPQF
QYGSKAYVKHPADIPDYMKLSFPEGFKWDRVMNFEDGGVVTVTQDSSLQDGEFIYKVKLRGTDFP
SDGPVMQKQTMGWEASTERMYPEDGALKGEVKMRLKLKDGGHYDAEVKTTYMAKKPVQLPGAYIT
DIKLDITSHNEDYTIVEQYERAEGRHSTGA

>33
MASSEDVIKEFMRFKVCMEGSVNGHEFEIEGEGEGRPYEGTQAAKLKVTKGGPLPFAWDILSPQF
QYGSKAYVKHPADIPDYMKLSFPEGFKWDRVMKFEDGGVVTVTQDSSLQDGVFIYKVKLRGTDFP
SDGPVMQKQTMGWEASIERMYPEDGLLKGEAKMRLKLKNGGHYDAEVKTTYMAKKPVQLPGAYIT
DIKLDITSHNEDYTIVEQYERAEGRHSTGA

>34
MASSEDVIKEFMRFKVYLEGSVNGHEFEIEGEGEGRPYEGTQVAKLKVTKGGPLPFAWDILSPQF
QYGSKAYVKHPADIPDYMKLSFPEGFKWDRVMNFEDGGVVTVTQDTSLQDGEFICKVKLRGTDFP
SEGPVMQKQTMGWEASTERMYPEDGALKGEDKMRLRLKDGGHYDADVKTTYMAKKPVQLPGAYIT
DIKLDITSHNEDYTIVEQYERAEGRHSTGA

>35
MASSEDVIKEFMRFKVYLEGSVNGHEFEIEGEGEGRPYEGTQVAKLKVTKGGPLPFAWDILSPQF
QYGSKAYVKHPADIPDYMKLSFPEGFKWDRVMNFEDGGVVTVTQDTSLQDGEFIYKVKLRGTDFP
SEGPVMQKQTMGWEASTERMYPEDGALKGEDKMRLRLKDGGHYEADVKTTYMAKKPVQLPGAYIT
DIKLDITSHNEDYTIVEQYERAEGRHSTGA

>36
MASSEDVIKEFMRFKVYLEGSVNGHEFEIEGEGEGRPYEGTQVAKLRVTKGGPLPFAWDILSPQF
QYGSKAYVKHPADIPDYMKLSFPEGFKWDRVMNFEDGGVVTVTQDTSLQDGEFICKVKLRGTDFP
SEGPVMQKQTMGWEASTERMYPEDGVLKGEDKMRLRLKDGGHYDAYVKTTYMAKKPVQLPGAYIT
DIKLDVTSHNEDYTIVEQYERAEGRHSTGA

>37
MASSEDVIKEFMRFKVSMEGSVNGHEFEIEGEGEGRPYEGTQTAKLKVTKGGPLPFAWDILSPQF
QYGSKAYVKHPADIPDYLKLSFPEGFKWERVMNFEDGGVVTVTQDSTLQDGVLIYKVKLRGTNFP
SDGPVMQKKTRGWEASTERMYPEDGVLKGEIKMRLKLKDGSHYEAEVKTTYMAKKPVQLPGAYKT
DIKLDITSHNEDYTIVEQYERAEGRHSTGA