








## Exploiting genetic diversity and gene synthesis to identify superior nitrogenase NifH protein variants to engineer N<sub>2</sub>-fixation in plants

Xi Jiang <sup>1,2</sup>, Lucía Payá-Tormo <sup>1,2</sup>, Diana Coroian<sup>1</sup>, Inés García-Rubio <sup>3</sup>, Rocío Castellanos-Rueda <sup>1,4</sup>, Álvaro Eserverri <sup>1,2</sup>, Gema López-Torrejón<sup>1,2</sup>, Stefan Burén <sup>1</sup>✉ & Luis Manuel Rubio <sup>1,2</sup>✉

Engineering nitrogen fixation in eukaryotes requires high expression of functional nitrogenase structural proteins, a goal that has not yet been achieved. Here we build a knowledge-based library containing 32 nitrogenase *nifH* sequences from prokaryotes of diverse ecological niches and metabolic features and combine with rapid screening in tobacco to identify superior NifH variants for plant mitochondria expression. Three NifH variants outperform in tobacco mitochondria and are further tested in yeast. *Hydrogenobacter thermophilus* (Aquificae) NifH is isolated in large quantities from yeast mitochondria and fulfills NifH protein requirements for efficient N<sub>2</sub> fixation, including electron transfer for substrate reduction, P-cluster maturation, and FeMo-co biosynthesis. *H. thermophilus* NifH expressed in tobacco leaves shows lower nitrogenase activity than that from yeast. However, transfer of [Fe<sub>4</sub>S<sub>4</sub>] clusters from NifU to NifH in vitro increases 10-fold the activity of the tobacco-isolated NifH, revealing that plant mitochondria [Fe-S] cluster availability constitutes a bottleneck to engineer plant nitrogenases.

<sup>1</sup>Centro de Biotecnología y Genómica de Plantas, Universidad Politécnica de Madrid, Instituto Nacional de Investigación y Tecnología Agraria y Alimentaria, Pozuelo de Alarcón, 28223 Madrid, Spain. <sup>2</sup>Departamento de Biotecnología-Biología Vegetal, Escuela Técnica Superior de Ingeniería Agronómica, Alimentaria y de Biosistemas, Universidad Politécnica de Madrid, 28040 Madrid, Spain. <sup>3</sup>Centro Universitario de la Defensa, Ctra. de Huesca s/n, 50090 Zaragoza, Spain. <sup>4</sup>Department of Biosystems Science and Engineering, ETH Zürich, 4058 Basel, Switzerland. ✉email: [stefan.buren@upm.es](mailto:stefan.buren@upm.es); [lm.rubio@upm.es](mailto:lm.rubio@upm.es)

Nitrogen (N) fertilizers used to increase crop productivity in intensive agriculture practices pollute groundwater and release greenhouse gasses<sup>1</sup>. On the other hand, subsistence agriculture practices including poor N fertilization produce low and inconsistent yields causing malnutrition and poverty<sup>2,3</sup>. There is large interest in engineering cereal crop varieties capable of acquiring their own N<sup>4</sup>. One approach to this outcome relies on functional expression of a nitrogenase enzyme by the cereal plant<sup>5</sup>. Nitrogenases are prokaryotic, O<sub>2</sub>-sensitive, two-component metalloproteins that convert inert N<sub>2</sub> into biologically useful NH<sub>3</sub><sup>6–8</sup>. The most efficient and widespread variant, the molybdenum nitrogenase, is composed of an Fe protein (*nifH*-encoded) and a MoFe protein (encoded by *nifD* and *nifK*). The Fe protein (NifH) donates electrons to the MoFe protein (NifDK) that in turn reduces N<sub>2</sub>. Nascent NifH and NifDK polypeptides need to acquire proper quaternary structure and to receive metal clusters, one [Fe<sub>4</sub>S<sub>4</sub>] cluster per NifH homodimer and two pairs of P-cluster and FeMo-co per NifDK heterotetramer, for functionality. We have recently reviewed the mechanisms and genetic requirements to assemble these cofactors and to mature NifH and NifDK into active Mo nitrogenase<sup>9</sup>. The large number of nitrogen fixation (*nif*) genes involved, and the sensitivity of most of the protein products towards O<sub>2</sub>, makes nitrogenase engineering a daunting task with issues that need to be solved stepwise.

To date, functional NifH, NifU, and NifB have been purified from mitochondria of aerobically cultured *Saccharomyces cerevisiae* cells<sup>10,11</sup>, while active NifU and NifH were isolated from chloroplasts of *Nicotiana benthamiana* at the end of the dark period<sup>12</sup>. Also, the reported low stability of the NifD protein<sup>13</sup> has now been improved in two recent studies that identified key residues in the NifD sequence as susceptible to cleavage upon mitochondria import<sup>14,15</sup>. Notwithstanding these achievements, detailed analysis of yeast mitochondria-targeted *Azotobacter vinelandii* NifH has been hampered by low protein solubility resulting in suboptimal yields<sup>10</sup>. Accumulation of mostly insoluble NifH was also reported when *Klebsiella oxytoca* NifH was targeted to the tobacco mitochondria<sup>16</sup>. The difficulty of expressing high levels of soluble and functional NifH in yeast and tobacco poses a major problem for eukaryotic nitrogenase engineering as it is the most abundant Nif protein during N<sub>2</sub> fixation<sup>17</sup>. The problem is exacerbated because, in addition to serving NifDK with electrons for substrate reduction, NifH is required to mature P-clusters onto NifDK and for the final steps of FeMo-co biosynthesis in complex with NifEN<sup>9</sup>. For these reasons it is essential to identify a NifH variant that is highly soluble and stable when expressed at very high levels in a plant cell, and that can perform all three NifH-dependent activities. One approach to achieve this outcome would be protein engineering of well-studied NifH from model diazotrophs (e.g., *A. vinelandii* or *K. oxytoca*) aimed to introduce sequences that improve stability in the mitochondria<sup>18</sup>. Protein engineering has been extensively employed to obtain glyphosate resistance<sup>19</sup>, another important trait for crops. Alternatively, mining of phylogenetically diverse *nifH* sources can be undertaken in order to find natural NifH proteins with superior properties, a strategy that was successful for NifB<sup>11</sup> and for increasing carotenoid levels in “Golden Rice”<sup>20</sup>.

Here, 32 distinct *nifH* genes were screened for expression level and solubility in mitochondria of *N. benthamiana*. The *nifM*, *nifU*, and *nifS* genes were co-expressed because their protein products are involved in NifH folding and in the biosynthesis and delivery of its [Fe<sub>4</sub>S<sub>4</sub>] cluster<sup>9</sup>. The *Hydrogenobacter thermophilus* NifH was identified as vastly superior to the *A. vinelandii* NifH in terms of expression levels, solubility, and functionality both in tobacco and yeast mitochondria. Mitochondria-targeted *H. thermophilus* NifH satisfied all functional and spectroscopic

requirements of a nitrogenase Fe protein when purified from yeast. The screening also pinpointed the plant mitochondria [Fe-S] cluster assembly as a bottleneck for further engineering.

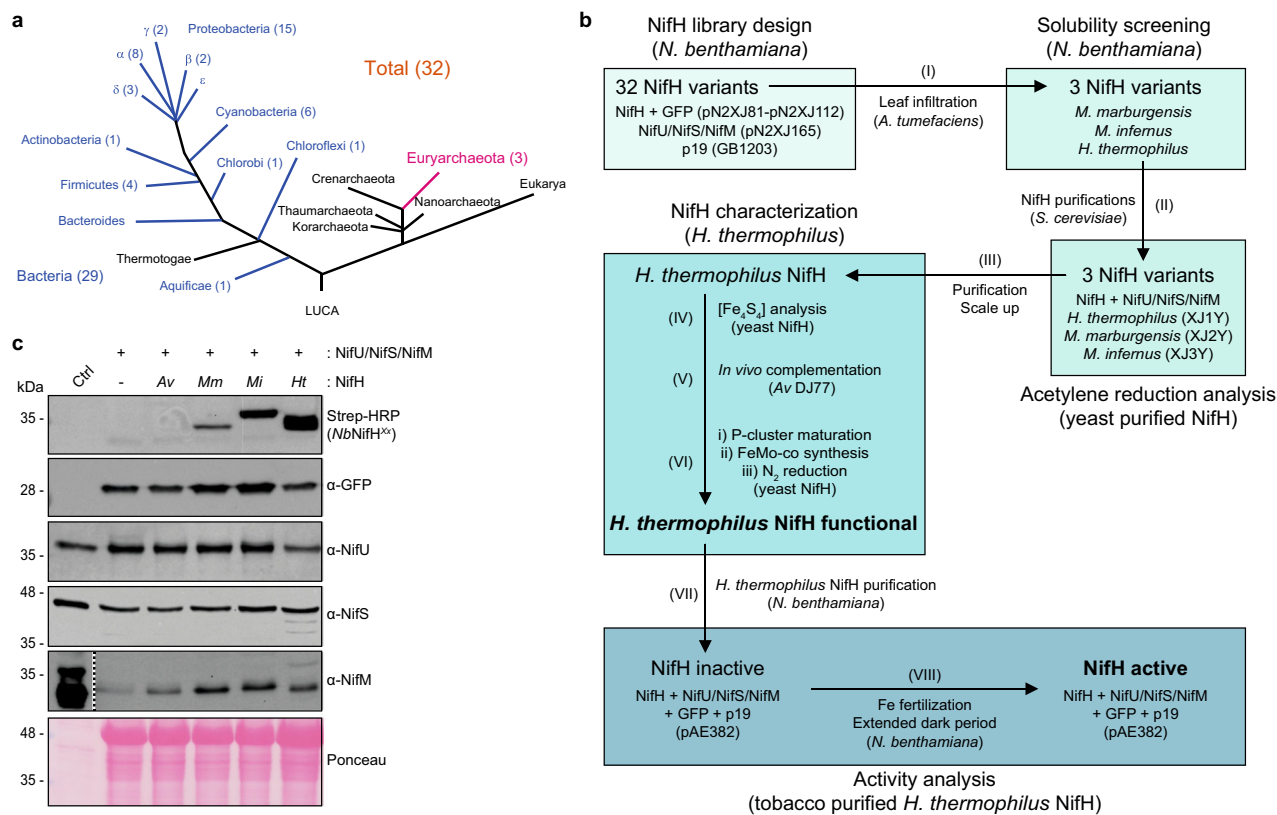
## Results

**Library design and strategy for expression of mitochondria-targeted NifH in *N. benthamiana*.** A library of 32 *nifH* sequences from phylogenetically diverse prokaryotes was designed considering one or several of the following criteria: (i) *nifH* genes found in confirmed diazotrophs; (ii) *nifH* genes from phototrophs or plant-associated bacteria; (iii) *nifH* genes from aerobic organisms; (iv) growth temperature of the *nifH* host; (v) *nifH* genes from archaeal representatives (Supplementary Data 1). Organized by phyla, the selection included genes from 1 Aquificae, 4 Firmicutes, 1 Actinobacteria, 15 Proteobacteria, 6 Cyanobacteria, 1 Chlorobi, 1 Chloroflexi, and 3 Euryarchaeota (Fig. 1a).

The workflow of this study is described in Fig. 1b. The gene sequences encoding the 32 NifH variants were cloned into plant vectors for *Agrobacterium tumefaciens* infiltration-mediated NifH expression in *N. benthamiana* leaves (Supplementary Table 1, see Methods section for details). The *nifH* sequences were codon-optimized for *S. cerevisiae* because codon-usage is similar to tobacco<sup>21</sup> and the workflow included downstream expression of tobacco-selected NifH variants in yeast for biochemical characterization. The genes were under control of the strong and constitutive E35S promoter. Amino-terminal COX4-TS extensions were added to NifH proteins. COX4 is the 29 amino acid transit peptide of the *S. cerevisiae* mitochondria protein cytochrome c oxidase subunit IV (MLSLRQSRIFKPTRTLCSSRYLLQKQP), whereas TS denotes the 28 amino acid Twin-Strep-Tag peptide (WSHPQFEKGGGGGGGGSSAWSHQPFEK)<sup>22</sup>. COX4 targeted NifH proteins to the mitochondria matrix and TS was used to enable variant-independent immunoblot detection of NifH and to facilitate its purification. Importantly, the TS-tag has been shown to not significantly affect NifH functionality<sup>12</sup>. COX4-TS-NifH variants are hereafter denoted as *NbNifH*<sup>Xx</sup> where *Nb* stands for the host *N. benthamiana*, <sup>Xx</sup> denotes variants collectively, and other superscripts indicate the species from which NifH sequence was obtained. Vectors with *NbNifH*<sup>Xx</sup> constructs additionally contained a transcriptional unit for expression of the green fluorescent protein (GFP) that was used as indicator of successful leaf infiltration (Supplementary Table 1).

An auxiliary vector was constructed to co-express *A. vinelandii* *nifM*, *nifU*, and *nifS* and target their protein products to mitochondria via N-terminal SU9 extensions. Similar to COX4, the mitochondrial presequence of subunit 9 of the *Neurospora crassa* F<sub>0</sub>-ATPase<sup>23</sup> (SU9) has been shown to deliver Nif proteins to *N. benthamiana* mitochondria<sup>24</sup>. NifU and NifS assemble [Fe-S] clusters destined for Nif proteins in *A. vinelandii*<sup>25</sup>. While not essential for expression of functional NifH<sup>Av</sup> in *S. cerevisiae* mitochondria<sup>10</sup> they were required to generate high amounts of active NifB in yeast<sup>11</sup>. As we aimed to identify NifH variants accumulating at higher levels than NifH<sup>Av</sup>, NifU<sup>Av</sup> and NifS<sup>Av</sup> were included in this study. In *A. vinelandii* and other well-studied diazotrophs NifM is involved in NifH folding or dimerization prior [Fe<sub>4</sub>S<sub>4</sub>] cluster acquisition<sup>9,26</sup>. Despite *nifM* not being present in organisms of some selected *nifH* variants (Supplementary Data 1), this gene was always included in infiltration experiments for consistency.

**Identification of NifH proteins suitable for expression in *N. benthamiana*.** *N. benthamiana* leaves were co-infiltrated with a 1:1:1 mixture of three distinct *A. tumefaciens* cultures for expression of, respectively, one *NbNifH*<sup>Xx</sup> variant plus GFP, the auxiliary proteins *NbNifM*<sup>Av</sup>, *NbNifU*<sup>Av</sup>, and *NbNifS*<sup>Av</sup>, and the



**Fig. 1** NifH library design and screening. **a** Phylogenetic tree indicating the distribution and number of the tested NifH variants among bacterial phyla (blue) and archaeal lineage (magenta). **b** Experimental workflow of this study. **c** Immunoblots of *N. benthamiana* soluble protein extracts developed with antibodies against TS, GFP, NifU, NifS, and NifM (step I of panel **b**). GFP detection was used as control to normalize effectiveness of infiltration experiments. Ponceau staining panel is shown as loading and membrane transfer control. Dotted line indicates different exposures of the same membrane. Cell-free extracts of *A. vinelandii* DJ (for NifU and NifS) or *E. coli* Rosetta (DE3) overexpressing NifM<sup>Av</sup> (for NifM, as NifM expression levels in *A. vinelandii* is low) were used as size controls. Uncropped immunoblots are shown in Supplementary Fig. 8.

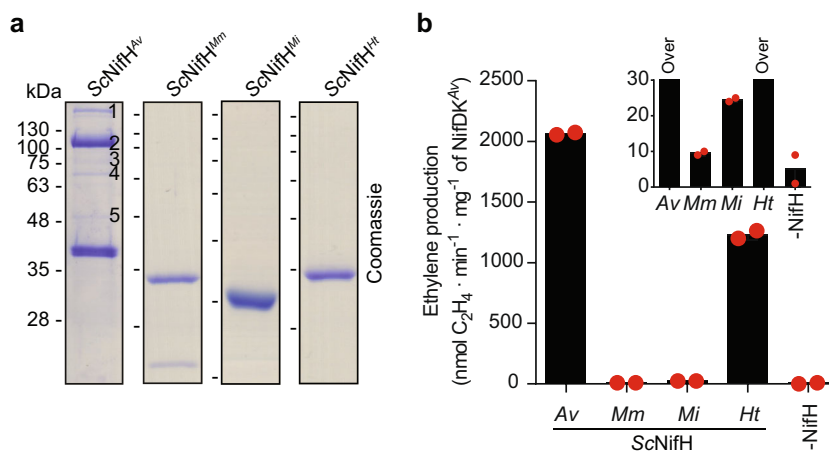
RNA silencing suppressor p19 to enhance the *nif* transgene expression (Fig. 1b)<sup>27</sup>. Protein extracts were prepared from the *N. benthamiana* leaves three days after infiltration and analyzed for accumulation of soluble NbNifH<sup>Xx</sup> using antibodies recognizing the TS-tag. Only two NifH variants were consistently detected among experiments (Fig. 1c, Supplementary Fig. 1a), namely those originating from *Methanocaldococcus infernus* (NbNifH<sup>Mi</sup>) and *Hydrogenobacter thermophilus* (NbNifH<sup>Ht</sup>). A third NifH variant from *Methanothermobacter marburgensis* (NbNifH<sup>Mm</sup>) was detected at low levels at one occasion. In contrast, analysis of total extracts prepared from the infiltrated tobacco leaves showed that, although accumulation levels of the NbNifH<sup>Xx</sup> proteins varied significantly, 25 of the 32 variants could be detected (Supplementary Fig. 1b). Only NbNifH expression of variants from *Bradyrhizobium japonicum*, *Rhizobium leguminosarum* bv. *trifolii*, *Herbaspirillum seropedicae*, *Gloeotheca* sp. KO68DGA, *Rhodopseudomonas palustris*, *Methanothermobacter thermoautotrophicus*, and *Frankia* sp. (strain FaC1) could not be demonstrated. Sequence alignments and 3D-modeling of NifH<sup>Mi</sup>, NifH<sup>Ht</sup>, and NifH<sup>Mm</sup> are shown in Supplementary Fig. 2. The 3D-models did not reveal any specific feature that would explain their superior accumulation as soluble protein in tobacco mitochondria, but all three proteins originate from thermophilic organisms (Supplementary Data 1) which could possibly explain their stability and solubility.

#### Activity of NifH variants isolated from mitochondria of aerobically cultured *S. cerevisiae*. *N. benthamiana* screening-

identified variants and NifH<sup>Av</sup> were expressed in *S. cerevisiae* and purified by Strep-tag affinity chromatography (STAC) to evaluate functionality when targeted to mitochondria. For this, genes encoding COX4-TS-NifH constructs were transferred to expression vectors together with *su9-nifM<sup>Av</sup>*, *su9-nifU<sup>Av</sup>*, and *su9-nifS<sup>Av</sup>* under the control of galactose-inducible GAL1 or GAL10 promoters (Supplementary Table 2, Supplementary Fig. 3a). These COX4-TS-NifH variants expressed in aerobic *S. cerevisiae* cultures are hereafter denoted ScNifH<sup>Mm</sup>, ScNifH<sup>Mi</sup>, ScNifH<sup>Ht</sup>, and ScNifH<sup>Av</sup> (ScNifH<sup>Xx</sup> collectively).

While ScNifH<sup>Mm</sup>, ScNifH<sup>Mi</sup>, and ScNifH<sup>Ht</sup> were purified to near homogeneity (Fig. 2a), SDS-PAGE analysis of ScNifH<sup>Av</sup> showed additional slower migrating co-eluting proteins. Mass spectrometry confirmed that these were contaminants (Fig. 2a). ScNifH<sup>Av</sup> solubility was low and much protein was lost to the pellet fraction when preparing the soluble cell-free extract (CFE) explaining its poor purification yield (about 11 mg per kg of *S. cerevisiae* cells) (Supplementary Fig. 3b–e, Supplementary Table 3). The yield of ScNifH<sup>Mm</sup> was also relatively low, in line with the inferior result in the *N. benthamiana* screening. In contrast, the yields of ScNifH<sup>Mi</sup> and ScNifH<sup>Ht</sup> were ca. 20 times higher. Iron (Fe) quantification of purified samples was variable but indicated that ScNifH<sup>Ht</sup> was isolated largely as holo-protein containing one [Fe<sub>4</sub>S<sub>4</sub>] cluster per dimer (Supplementary Table 3). Consistently, immunoblot analysis showed that ScNifH<sup>Ht</sup>, ScNifM<sup>Av</sup>, ScNifU<sup>Av</sup>, and ScNifS<sup>Av</sup> had been efficiently targeted to the mitochondria (Supplementary Fig. 4).

Activities of purified ScNifH<sup>Xx</sup> variants were determined in vitro using the acetylene reduction assay (ARA) and compared



**Fig. 2** Activity of selected NifH variants purified from mitochondria of aerobically cultured *S. cerevisiae* engineered strains. **a** STAC-purified ScNifH<sup>Av</sup> (strain XJ4Y), ScNifH<sup>Mm</sup> (strain XJ2Y), ScNifH<sup>Mi</sup> (strain XJ3Y), and ScNifH<sup>Ht</sup> (strain XJ1Y) proteins. The low solubility of ScNifH<sup>Av</sup> promoted binding of contaminating proteins (as column was not saturated with ScNifH<sup>Av</sup>) that were identified by peptide mass fingerprinting as: 1) Acc1p (N1P4Q3), 2 and 3) pyruvate carboxylase (N1P377), and 4) HSP70 (D2J4C2). No identification was possible for band number 5. Uncropped gels are shown in Supplementary Fig. 9. **b** ARA of STAC-purified ScNifH<sup>Xx</sup> variants. Activity using NifH<sup>Av</sup> and NifDK<sup>Av</sup> (positive control) was  $2406 \pm 53$  units (nmol ethylene formed per min and mg of NifDK<sup>Av</sup>). Data represent mean values ( $n = 2$  technical replicates).

to that of NifH purified from *A. vinelandii* (denoted NifH<sup>Av</sup>). In all cases NifDK purified from *A. vinelandii* (denoted NifDK<sup>Av</sup>) was used as MoFe protein component. ScNifH<sup>Av</sup> activity was 85% of NifH<sup>Av</sup> (Fig. 2b), supporting previous observations that STAC is suitable for purification of metal-cluster containing Nif proteins expressed in yeast<sup>11,28</sup>. ScNifH<sup>Ht</sup> specific activity was about half of ScNifH<sup>Av</sup>, while ScNifH<sup>Mm</sup> and ScNifH<sup>Mi</sup> showed very low activities (Fig. 2b). The assay did not determine whether lower activities were due to NifH variant defects, or to incompatibility with ScNifUS<sup>Av</sup> in vivo (resulting in apo-NifH protein with low [Fe<sub>4</sub>S<sub>4</sub>] cluster occupancy) or NifDK<sup>Av</sup> in vitro (resulting in poor electron donation). Reconstitution of ScNifH<sup>Mm</sup> [Fe<sub>4</sub>S<sub>4</sub>] clusters in vitro by either mixing with Fe, L-cysteine, DTT, and EcNifS<sup>Av</sup> (direct reconstitution) or by incubating with [Fe<sub>4</sub>S<sub>4</sub>] cluster-loaded EcNifU<sup>Av</sup> (NifU-mediated reconstitution) did not activate the protein (Supplementary Fig. 5), indicating that this NifH variant is not compatible with NifDK<sup>Av</sup>. In contrast, ScNifH<sup>Mi</sup> was activated to some extent by NifU<sup>Av</sup>, and further by direct reconstitution, indicating that the *A. vinelandii* NifUS machinery is not optimal for NifH<sup>Mi</sup> (Supplementary Fig. 5). However, activities were very low compared to the as-isolated ScNifH<sup>Ht</sup> protein (Fig. 2b). This could be explained by NifH<sup>Ht</sup> harboring more of the conserved amino acid residues known to be important for the interaction with NifDK<sup>Av</sup> (Supplementary Fig. 2).

Importantly, soluble accumulation of ScNifH<sup>Ht</sup> in mitochondria was 20-fold higher than ScNifH<sup>Av</sup> (Supplementary Table 3), which translates into at least 10-fold higher in vivo activity and fulfills NifH quantity requirements for nitrogenase engineering. Thus, ScNifH<sup>Ht</sup> was further characterized.

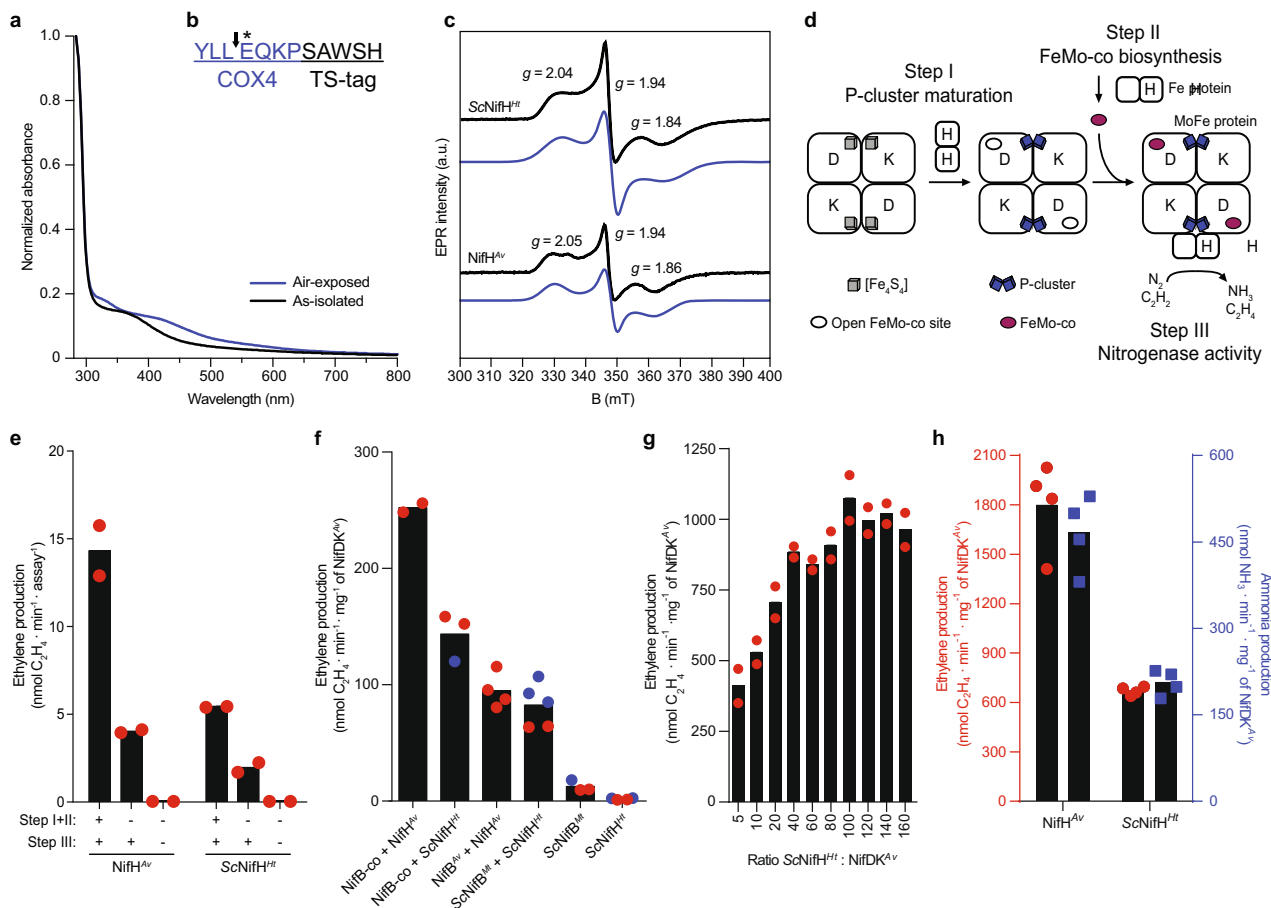
**ScNifH<sup>Ht</sup> exhibits NifH-characteristic spectroscopic signals and is functional in vivo.** Purified ScNifH<sup>Ht</sup> protein presented ultraviolet-visible (UV-vis) absorption spectra typical of O<sub>2</sub>-sensitive [Fe-S] cluster-containing proteins (Fig. 3a). Amino-terminal sequencing revealed that amino acid residues EQKP remained after COX4 processing (Fig. 3b), where conversion of glutamine (Q) to glutamic acid (E) could be due to deamination performed by the mitochondrial matrix N-terminal amidase NTA1<sup>29</sup>. Electron paramagnetic resonance (EPR) confirmed that ScNifH<sup>Ht</sup> protein contained an [Fe<sub>4</sub>S<sub>4</sub>] cluster with similar signal

intensity and *g*-values as NifH<sup>Av</sup> (Fig. 3c), suggestive of successful maturation into functional Fe protein.

The NifH variant chosen to engineer N<sub>2</sub>-fixing plants must perform P-cluster maturation and FeMo-co biosynthesis in addition to serve as electron donor for substrate reduction. We therefore tested whether *H. thermophilus* NifH could revert the Nif<sup>-</sup> phenotype of *A. vinelandii* DJ77 ( $\Delta$ nifH strain)<sup>30</sup>. For this, *ts-nifH<sup>Ht</sup>* was introduced by transformation into DJ77 and the resulting strain UW481 was tested for diazotrophic growth and in vivo acetylene reduction activity. UW481 showed diazotrophic growth both in solid and liquid media (Supplementary Fig. 6a, b), and immunoblot analysis demonstrated sustained AvNifH<sup>Ht</sup> expression and acetylene reducing activity indicative of active nitrogenase (Supplementary Fig. 6c, d). These data strongly indicate that NifH<sup>Ht</sup> can replace the functions of native *A. vinelandii* NifH to some extent, which requires productive interactions with at least apo-NifDK<sup>Av</sup>, NifDK<sup>Av</sup>, and NifEN<sup>Av</sup> proteins.

**ScNifH<sup>Ht</sup> is active in substrate reduction, P-cluster formation and FeMo-co synthesis.** Each individual NifH-dependent activity was then analyzed in vitro using pure ScNifH<sup>Ht</sup> preparations (Fig. 3d). P-cluster maturation was determined by supplementing CFE of *A. vinelandii* DJ77 ( $\Delta$ nifH) with ScNifH<sup>Ht</sup>. The DJ77 extract is devoid of FeMo-co and contains inactive apo-NifDK<sup>Av</sup> with immature P-clusters. The P-cluster maturation assay using DJ77 CFE relies on positive outcomes of three distinct activities performed in two sequential reactions (Fig. 3d). In the first reaction (Step I + II) pure NifH and FeMo-co are added to DJ77 CFE resulting in NifH-dependent reductive coupling of the two [Fe<sub>4</sub>S<sub>4</sub>] P-cluster precursors to form mature P-clusters (Step I), followed by FeMo-co insertion into P-cluster containing apo-NifDK<sup>Av</sup> to generate active NifDK<sup>Av</sup> (Step II) (Fig. 3d). Tetra-thiomolybdate is then added to prevent further FeMo-co insertion, separating the maturation (Step I + II) and activity (Step III) reactions. Activation of DJ77 apo-NifDK<sup>Av</sup> by ScNifH<sup>Ht</sup> demonstrated its P-cluster maturation activity (Fig. 3e).

In vitro FeMo-co synthesis (Fig. 3d, Step II)<sup>9</sup> was determined by combining purified preparations of ScNifH<sup>Ht</sup>, apo-NifDK<sup>Av</sup> containing P-clusters but devoid of FeMo-co<sup>31</sup>, apo-NifEN<sup>Av</sup> containing permanent [Fe<sub>4</sub>S<sub>4</sub>] clusters but lacking FeMo-co precursor<sup>32</sup>, Mo, homocitrate, and either the FeMo-co precursor



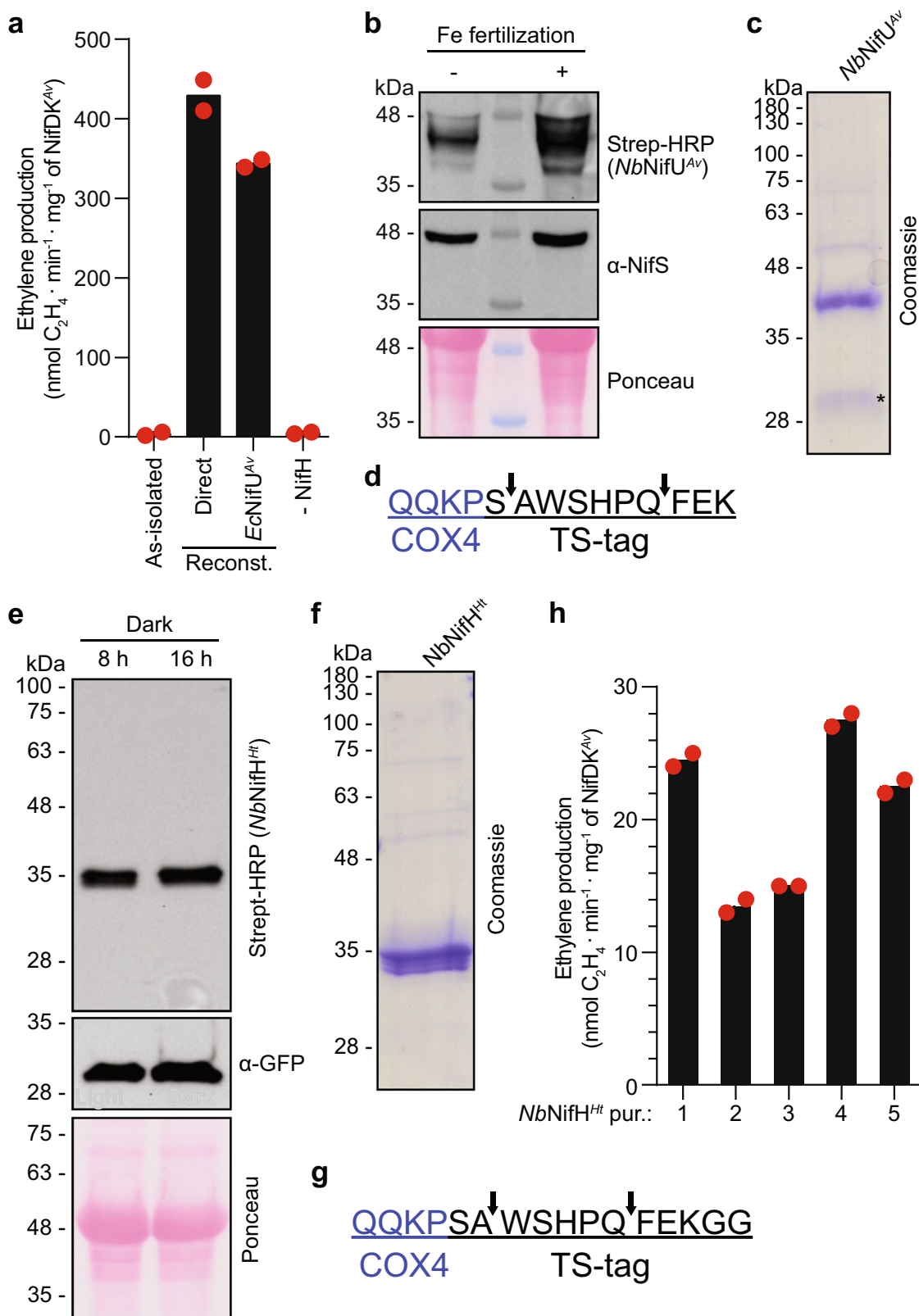
**Fig. 3 Characterization of ScNifH<sup>Ht</sup>.** **a** UV-vis absorption spectra of as-isolated and air-exposed ScNifH<sup>Ht</sup>. **b** Processing site (black arrow) of COX4 mitochondria targeting signal (blue) as determined by N-terminal sequencing of ScNifH<sup>Ht</sup>. The conversion of Q to E (marked by \*) could be due to deamination processes. **c** EPR signal of as-isolated ScNifH<sup>Ht</sup> (190  $\mu$ M) compared to NifH<sup>Av</sup> (71.2  $\mu$ M). Experimental data (black lines) and simulations (blue lines) of each protein are shown together with  $g$  values. **d** Schematic representation of NifH-dependent activities tested in **e-h**. Figure adapted from Burén et al.<sup>9</sup>. Copyright 2020 ACS under CC BY 4.0 <http://creativecommons.org/licenses/by/4.0>. **e** In vitro P-cluster maturation of apo-NifDK<sup>Av</sup> present in CFE of *A. vinelandii* DJ77 ( $\Delta$ NifH) after addition of ScNifH<sup>Ht</sup> or NifH<sup>Av</sup> as indicated ( $\pm$ ). Tetrathiomolybdate was added (following step I + II) to inhibit further apo-NifDK<sup>Av</sup> activation during the ARA (step III). Data represent mean values ( $n = 2$  technical replicates). **f** NifB-co and NifB-dependent in vitro FeMo-co synthesis using ScNifH<sup>Ht</sup> or NifH<sup>Av</sup>. Tetrathiomolybdate was added (following step II) to inhibit further apo-NifDK<sup>Av</sup> activation during the ARA (step III). A 20:1 molar ratio of ScNifH<sup>Ht</sup> to NifDK<sup>Av</sup> was used in the ARA (step III). Data represent mean values ( $n = 2$  technical replicates (NifB-co + NifH<sup>Av</sup>),  $n = 3$  technical replicates (NifB-co + ScNifH<sup>Ht</sup>),  $n = 4$  technical replicates (NifB<sup>Av</sup> + NifH<sup>Av</sup>),  $n = 5$  technical replicates (ScNifB<sup>Mt</sup> + ScNifH<sup>Ht</sup>),  $n = 3$  technical replicates (ScNifB<sup>Mt</sup>),  $n = 4$  technical replicates (ScNifH<sup>Ht</sup>)). Blue and red dots correspond to two independent experiments. **g** Titration of NifDK<sup>Av</sup> activity with ScNifH<sup>Ht</sup>. Positive control reactions performed with NifH<sup>Av</sup> and NifDK<sup>Av</sup> at 40:1 molar ratio gave  $1692 \pm 4$  units (nmol ethylene formed per min and mg of NifDK<sup>Av</sup>). Reactions lacking NifH (negative control) gave  $2.5 \pm 0.8$  units. Data represent mean values ( $n = 2$  technical replicates). **h** ARA (red dots, left y-axis) and N<sub>2</sub>-reduction assay (blue squares, right y-axis) using ScNifH<sup>Ht</sup> and NifDK<sup>Av</sup> (step III). NifH<sup>Av</sup> was used as control. Data represent mean values ( $n = 4$  technical replicates).

NifB-co bound to the carrier protein NifX<sup>Av33</sup> or NifB protein supplemented with Fe and S<sup>34</sup>. As for the P-cluster maturation assay, tetrathiomolybdate was added before the ARA (Fig. 3d, Step III). Figure 3f shows that ScNifH<sup>Ht</sup> supported FeMo-co synthesis in vitro. Importantly, ScNifH<sup>Ht</sup> and ScNifB<sup>Mt</sup> (*Methanothermobacter thermautotrophicus* NifB isolated from *S. cerevisiae*)<sup>11</sup> acted together in the NifB-dependent in vitro FeMo-co synthesis assay in which NifB-co was concomitantly synthesized by ScNifB<sup>Mt</sup> rather than added in purified form. This result proved compatibility of two essential proteins for N<sub>2</sub> fixation, ScNifH<sup>Ht</sup> and ScNifB<sup>Mt</sup>, when produced in yeast mitochondria. It also showed interspecies compatibility with NifDK<sup>Av</sup> and NifE<sup>Av</sup>, altogether constituting the conserved biochemical core of nitrogenase.

ScNifH<sup>Ht</sup> activity in substrate reduction was demonstrated by the ARA and by reduction of N<sub>2</sub> into NH<sub>3</sub>. ARA titration was

carried out with a fixed quantity of NifDK<sup>Av</sup> and increasing amounts of ScNifH<sup>Ht</sup>. Maximum NifDK<sup>Av</sup> activity was achieved at molar ScNifH<sup>Ht</sup> to NifDK<sup>Av</sup> ratios larger than 40 (Fig. 3g), similar to reactions with the natural counterpart NifH<sup>Av</sup><sup>35</sup>. This result suggests that the maximum activity that can be achieved combining ScNifH<sup>Ht</sup> with NifDK<sup>Av</sup> is 1000 units (i.e., half of the activity with NifH<sup>Av</sup>). In addition, ScNifH<sup>Ht</sup> supported N<sub>2</sub> reduction into NH<sub>3</sub> by NifDK<sup>Av</sup>. Importantly, the ratio of NH<sub>3</sub> to ethylene produced by NifDK<sup>Av</sup> was similar independently of using NifH<sup>Av</sup> or ScNifH<sup>Ht</sup> (Fig. 3h).

**As-isolated NbNifH<sup>Ht</sup> was inactive but could be activated by [Fe<sub>4</sub>S<sub>4</sub>] cluster reconstitution.** NbNifH<sup>Ht</sup> was purified from *A. tumefaciens*-infiltrated leaves of *N. benthamiana*. Plants were grown under long-day conditions (16 h light/8 h dark) and leaves were processed at the end of the dark period. Genes encoding



*NbnifH<sup>Ht</sup>*, *NbnifM<sup>Av</sup>*, *NbnifU<sup>Av</sup>*, and *NbnifS<sup>Av</sup>* (together with p19 and GFP) were piled up in a single plant-expression vector for co-expression (Methods section and Supplementary Table 1). Purified *NbnifH<sup>Ht</sup>* did not exhibit brown color of [Fe-S] clusters and was inactive in the ARA when combined with *NifDK<sup>Av</sup>* (Fig. 4a). Therefore, we reconstituted *NbnifH<sup>Ht</sup>* [Fe<sub>4</sub>S<sub>4</sub>] cluster in vitro either by mixing with Fe, L-cysteine, DTT, and *EcNifS<sup>Av</sup>*

(direct reconstitution), or by incubating with [Fe<sub>4</sub>S<sub>4</sub>] cluster-loaded *EcNifU<sup>Av</sup>* (*NifU*-mediated reconstitution). Both methods activated the *NbnifH<sup>Ht</sup>* as determined by the ARA (Fig. 4a), demonstrating that the protein was correctly folded but lacked its [Fe<sub>4</sub>S<sub>4</sub>] cluster. This result suggested that insertion and/or stability of *NbnifH<sup>Ht</sup>* [Fe<sub>4</sub>S<sub>4</sub>] cluster was poor in mitochondria of leaves.

**Fig. 4 Characterization of the *NbNifH<sup>Ht</sup>* protein.** **a** Activation of as-isolated *NbNifH<sup>Ht</sup>* protein with  $[\text{Fe}_4\text{S}_4]$  clusters either by direct chemical synthesis or by *EcNifU<sup>Av</sup>*-mediated reconstitution. Activity using *NifH<sup>Av</sup>* (positive control) was  $1773 \pm 10$  units (nmol ethylene formed per min and mg of *NifDK<sup>Av</sup>*). Data represent mean values ( $n = 2$  technical replicates). **b** Immunoblots showing the effect of Fe fertilization on *NbNifU<sup>Av</sup>* and *NbNifS<sup>Av</sup>* total protein expression. *NifS* antibody was probed on the same membrane after incubation with Streptactin-HRP. **c** *NbNifU<sup>Av</sup>* protein purified from tobacco plants. The lower band (marked by \*) indicates a faster migrating *NbNifU<sup>Av</sup>* polypeptide. **d** Processing sites (black arrows) of the COX4 mitochondria targeting signal (blue) as determined by N-terminal sequencing of the full-length and the faster migrating *NbNifU<sup>Av</sup>*. **e** Immunoblots showing accumulation of *NbNifH<sup>Ht</sup>* at the end of 8 h or 16 h night (dark period). **f** *NbNifH<sup>Ht</sup>* protein purified from tobacco leaves. **g** Processing sites (black arrows) of COX4 signal (blue) as determined by N-terminal sequencing of *NbNifH<sup>Ht</sup>*. **h** ARA of five independent *NbNifH<sup>Ht</sup>* STAC-purifications (1–5). Measured activities using *NifH<sup>Av</sup>* (positive controls) and without *NifH* (negative controls) were, respectively,  $2406 \pm 53$  and  $4.6 \pm 1.3$  (purification 1),  $1773 \pm 10$  and  $3.2 \pm 2.2$  (purifications 2 and 3), and  $1692 \pm 3.9$  and  $2.5 \pm 0.8$  (purifications 4 and 5). All activities are in nmol ethylene formed per min and mg of *NifDK<sup>Av</sup>*. Data represent mean values ( $n = 2$  technical replicates). Uncropped immunoblots and gels are shown in Supplementary Fig. 10.

**Fe fertilization of the soil increases soluble NifU in mitochondria of *N. benthamiana*.** One explanation for the low  $[\text{Fe}_4\text{S}_4]$  cluster content of *NbNifH<sup>Ht</sup>* could be insufficient Fe availability in the soil. We observed that accumulation of *NbNifU<sup>Av</sup>*, but not of *NbNifS<sup>Av</sup>*, increased when the water used to irrigate the *A. tumefaciens*-infiltrated plants was supplemented with Fe (Fig. 4b). Sulfur was not supplemented in soil as the infiltration solution contained  $\text{Mg}_2\text{SO}_4$ . Although Fe fertilization tripled the yield of STAC-isolated *NbNifU<sup>Av</sup>*, the average Fe content of 2 Fe atoms per protein was not affected (Supplementary Fig. 7a, b, Supplementary Table 4). This could be due to the loss of transient NifU [Fe-S] clusters during purification, and it is not a surprising outcome as isolation of *EcNifU<sup>Av</sup>* containing only the permanent  $[\text{Fe}_2\text{S}_2]$  clusters has been previously observed<sup>36</sup>. Immunoblots detected two differently migrating *NbNifU<sup>Av</sup>* species in purifications from tobacco leaves (Fig. 4c). Amino-terminal sequencing showed that both species were cleaved either one or seven amino acids into the TS-tag (Fig. 4d). As both *NbNifU<sup>Av</sup>* species showed the same N-termini processing, we concluded that the faster migrating polypeptide was truncated at the C-terminus.

**Extended dark period combined with Fe fertilization produced active *NbNifH<sup>Ht</sup>* in mitochondria of *N. benthamiana* leaves.** The soil of *N. benthamiana* plants expressing *NbNifH<sup>Ht</sup>* was fertilized with Fe to increase Fe availability. In addition, the dark period preceding leaf harvest was extended from 8 h to 16 h hypothesizing that longer darkness would lower intracellular  $\text{O}_2$  and stabilize *NbNifH<sup>Ht</sup>*  $[\text{Fe}_4\text{S}_4]$  cluster. Dark period extension did not increase *NbNifH<sup>Ht</sup>* accumulation (Fig. 4e) but allowed for isolation of active protein as shown below. About 6 mg of *NbNifH<sup>Ht</sup>* was consistently isolated per kg of *N. benthamiana* leaves (Fig. 4f, Supplementary Fig. 7c, Supplementary Table 5). Amino-terminal sequencing showed that mitochondria-targeted *NbNifH<sup>Ht</sup>* accumulated as two species (similar to *NbNifU<sup>Av</sup>*), one in which two amino acid residues from the TS-tag were removed with the COX4 signal and another that was processed five amino acid residues further into the TS-tag (Fig. 4g).

Functionality of *NbNifH<sup>Ht</sup>* isolated from leaves of Fe fertilized tobacco plants following 16 h of darkness was determined using ARA. *NbNifH<sup>Ht</sup>* preparations consistently showed activities but these were low compared to those of  $[\text{Fe}_4\text{S}_4]$  cluster-reconstituted *NbNifH<sup>Ht</sup>* (Fig. 4h). This result suggested that *NbNifH<sup>Ht</sup>* accumulated as two species in tobacco mitochondria, where inactive protein likely lacking  $[\text{Fe}_4\text{S}_4]$  cluster was more abundant than functional and  $[\text{Fe}_4\text{S}_4]$  cluster-containing *NbNifH<sup>Ht</sup>*. Consistently, the Fe content of purified *NbNifH<sup>Ht</sup>* preparations were below detection limit (Supplementary Table 5). Altogether the results indicate that while soluble *NbNifH<sup>Ht</sup>* accumulates in good quantity in mitochondria of *N. benthamiana* leaves, engineering of additional protein components or biosynthetic pathways will be required to improve  $[\text{Fe}_4\text{S}_4]$  cluster acquisition or stability.

## Discussion

The first study reporting production of active *NifH* in yeast proved that mitochondria is a suitable organelle for hosting  $\text{O}_2$ -sensitive *Nif* proteins under aerobic growth conditions<sup>10</sup>. Despite being a valid proof-of-concept, further developments with *A. vinelandii* *NifH* were limited by low yields as only a small portion was soluble in the mitochondrial matrix. Similar solubility issues were later reported for *K. oxytoca* *NifH* targeted to *N. benthamiana* mitochondria<sup>16</sup> and are confirmed in this study using immunoblot screening and STAC. Identifying the best possible *NifH* protein for eukaryotic (plant) expression was therefore of uttermost importance. *NifH* is the most abundant *Nif* protein required for  $\text{N}_2$  fixation in *A. vinelandii*<sup>17</sup>. Besides being the Fe protein component of Mo nitrogenase, *NifH* is essential to the assembly of both *NifDK* cofactors, namely the P-cluster and the FeMo-co<sup>9</sup>.

*NifH* proteins for nitrogenase engineering in plants should: (i) be stable and soluble at high levels in the mitochondrial matrix, and (ii) be compatible with the *NifDK* component from a well-studied model-diazotroph if their own *NifDK* components are not available in purified form. Compatibility is important when evaluating function of candidate *NifH* variants. In our case it meant that any selected *NifH<sup>Xx</sup>* should be compatible with *NifM<sup>Av</sup>* (if *NifH<sup>Xx</sup>* is not *NifM*-independent), *NifU<sup>Av</sup>* for maturation and  $[\text{Fe}_4\text{S}_4]$  cluster synthesis/insertion, and *NifDK<sup>Av</sup>* for nitrogenase activity measurements. We note that this requirement introduces a selection bias and that the screening could have overlooked *NifH* variants that were superior to that of *H. thermophilus* if combined with different *NifDK*.

The *NifH* variants tested in this study were selected from a curated dataset of hundreds of *NifH* sequences by favoring aerobic or plant-associated origins, to overcome the inherent  $\text{O}_2$ -sensitivity of *NifH*, and functionality at moderate temperatures. We also hypothesized that *NifH* variants from archaea could function better in a eukaryotic environment as this domain of life is believed to be more closely related to the Eukaryota<sup>37</sup>, and because our previous work expressing archaeal *NifB* variants in yeast had shown them to be superior to those of bacterial origin<sup>11</sup>.

We expected that most *NifH* variants would be partly soluble in tobacco mitochondria when expressed together with the accessory proteins *NifU<sup>Av</sup>*, *NifS<sup>Av</sup>*, and *NifM<sup>Av</sup>*. However, only *NifH* from *M. infernus* and *H. thermophilus* were consistently detected in soluble tobacco extracts, in addition to *M. marburgensis* that was occasionally detected at lower levels. Two of these *NifH* proteins originated from archaea and the third from a bacterium. One possibility could be that the *NifM<sup>Av</sup>* protein was not expressed at sufficient levels in the tobacco mitochondria and that only these three *NifH* variants did not require *NifM* for maturation. However, low levels of *NifM* expression appear to be enough for *NifH* maturation in *K. oxytoca*<sup>38,39</sup>. A more plausible explanation can be found in the thermophilic nature of *M. infernus*, *H. thermophilus*, and *M. marburgensis*. It has recently

been reported that the temperature inside respiring mitochondria of cultured human cells is around 50 °C<sup>40</sup>, even when the external medium is maintained at 38 °C. Whether the same drastic effect on temperature holds true for mitochondria of a leaf cell is not known to us, but it could explain in part the outcome of our NifH screening. None of the two highest expressed NifH proteins originated from proven diazotrophs. We are not aware of any study investigating diazotrophy in the archaeon *M. infernus*. However, NifB<sup>Mi</sup> cured the Nif<sup>-</sup> phenotype of an *A. vinelandii* *nifB* mutant strain<sup>41</sup> and, as NifB has no other known function than biosynthesis of nitrogenase active-site cofactors, it is likely that *M. infernus* is in fact a diazotroph. On the other hand, N<sub>2</sub>-fixation has been tested but not observed in *H. thermophilus* TK-6<sup>42</sup>. Interestingly only six NifH variants in our library originated from organisms having genes with high similarity to *A. vinelandii* *nifM*. Perhaps other prolyl isomerases could substitute for NifM in these organisms. Whether NifM (and the NifUS machinery) is required for maturation of the three selected NifH proteins (especially NifH<sup>Ht</sup>) in mitochondria will be investigated in future work.

Mitochondria-expressed ScNifH<sup>Ht</sup> was the only variant that supported relevant nitrogenase activity when combined with NifDK<sup>Av</sup>. Its activity corresponded to roughly half of that using ScNifH<sup>Av</sup> even if the ScNifH<sup>Ht</sup> to NifDK<sup>Av</sup> molar ratio was increased well above 40 normally used for ARA. Emerich and Burris showed that NifH proteins can function with NifDK from other organisms<sup>35</sup>, but this study only combined proteins from bacteria. An optimal growth temperature of 72 °C has been reported for *H. thermophilus* TK-6<sup>43</sup>, which could explain lower ScNifH<sup>Ht</sup> activity in substrate reduction assays. However, our prediction from this study and previous work on NifB is that suboptimal working temperature of Nif proteins from thermophiles is a price worth paying when engineering nitrogenase in eukaryotes, as solubility and stability of these variants is so much improved.

One observation of this study was that the specific activity of as-isolated NbNifH<sup>Ht</sup> protein was lower than ScNifH<sup>Ht</sup>. We think this was caused by poor [Fe<sub>4</sub>S<sub>4</sub>] cluster availability – and hence inefficient incorporation – or poor NifH [Fe<sub>4</sub>S<sub>4</sub>] cluster stability within the leaf cell mitochondria. In this context, it is not known how Fe fertilization increased accumulation of soluble NbNifU<sup>Av</sup>. More available Fe could increase mitochondria [Fe-S] clusters biosynthesis and [Fe<sub>2</sub>S<sub>2</sub>] cluster occupancy in NbNifU<sup>Av</sup> which, in turn, would provide stability to the protein. A compatibility issue between NbNifH<sup>Ht</sup> and NbNifU<sup>Av</sup> and NbNifS<sup>Av</sup> is unlikely since EcNifU<sup>Av</sup> could effectively activate NbNifH<sup>Ht</sup> in vitro. NbNifH<sup>Ht</sup> misfolding in mitochondria is also unlikely as it was efficiently activated by reconstitution of its [Fe<sub>4</sub>S<sub>4</sub>] cluster. It is however likely that protection by respiratory O<sub>2</sub>-consumption in leaf is lower than in yeast. NbNifH<sup>Ht</sup> exposure to O<sub>2</sub> during leaf processing is also a possibility making this a purely technical problem. While leaves were kept in liquid nitrogen and lysis and purification were performed inside an anaerobic glove box, it is difficult to completely rule out that some O<sub>2</sub> trapped within the leaf was released during tissue disruption.

In conclusion, this study shows that genetic diversity can be exploited to identify, from a very large pool of sequences, the most adequate Nif protein components to engineer a eukaryotic nitrogenase. Modular cloning techniques, gene synthesis with codon optimization, and other synthetic biology tools permit building multi-protein pathways with components of very diverse origin. In this case the NifH protein from *H. thermophilus* was identified as soluble in mitochondria of both *S. cerevisiae* and *N. benthamiana* accumulating at much higher levels than the *A. vinelandii* homologue. This example is relevant not only because the identified variant performed all three NifH-essential reactions,

namely P-cluster maturation, FeMo-co biosynthesis, and NifD-K<sup>Av</sup> reduction, but also because NifH<sup>Ht</sup> formed functional interspecies interactions with NifB, NifE, and NifDK proteins, altogether representing the four proteins constituting the core of diazotrophy.

## Methods

**Design, assembly, and cloning of the nifH library.** A curated dataset of diazotrophs<sup>41</sup> was used to collect *nifH* candidates and design the library. Genes encoding *nifH* variants were codon optimized for expression in *S. cerevisiae* and the sequence encoding *pE35S::cox4-twinstrep* was codon optimized for expression in tobacco (Supplementary Data 1). All genetic parts were optimized using the GeneOptimizer tool (ThermoFisher) and synthesized by ThermoFisher via the Engineering Nitrogen Symbiosis for Africa (ENSA) project. The *nifH* genes were synthesized and cloned into pMA cloning vector with *Bam*HI and *Bst*EII restriction sites flanking each gene. The *pE35S::cox4-twinstrep* sequence was flanked by *Hind*III and *Bgl*II restriction sites.

pGFPUSplus (plasmid #64401, Addgene) and the pMA vector containing *pE35S::cox4-twinstrep* were digested with *Hind*III and *Bgl*II and used to generate the parental vector pN2SB41, containing a *pE35S::cox4-twinstrep-gus-tNOS* transcriptional unit in which *gus* was flanked by *Bam*HI and *Bst*EII restriction sites. The parental vector pN2SB41 and all pMA vectors containing *nifH* variants were digested with *Bam*HI and *Bst*EII and used to generate vectors pN2XJ81-pN2XJ112 (Supplementary Table 1).

pGFPUSplus was used to generate vector pN2XJ165 containing transcriptional units for mitochondria-targeted accessory Nif proteins (*A. vinelandii* NifU, NifS, and NifM). The *su9-nifU<sup>Av</sup>* (AAAAGGATCCAATGGCCTCCACTCGTGTCTCG, AAAAAAGGTCACCTTAGACTTCCATTGGGCGTGTGCG) and *su9-nifS<sup>Av</sup>* (AACTAGTAGTGGCCTCCACTCGTGTCTCG, AAAAGAGCTCTTAACCATAGACAGGAGCAAAGGCTTACC) genes were amplified by PCR from the yeast vector pN2GLT4<sup>10</sup>. Amplification reactions added flanking *Bam*HI and *Bst*EII (for *su9-nifU*) or *Spe*I and *Sac*I (for *su9-nifS<sup>Av</sup>*) sites. The DNA fragment containing the *su9-nifM<sup>Av</sup>* sequence was created by overlapping PCR using primers introducing sequences homologous to those flanking the *Xho*I site of pGFPUSplus (ATTATG GAGAACTCGAGTTAACCATGTGCTAAGTTTCC, TACAAATCTATCTCTCTCGAGATGGCCTCCACTCGTG, CTCTCTGAGGCCATGGAAGAGTAGGCG CTCCTCTG, CGCGCCTACTCTTCCATGGCCTCAGAAAGATTAGCTGATG). pGFPUSplus was first digested with *Bgl*II and *Bst*EII to insert *su9-nifU<sup>Av</sup>*, then with *Xba*I and *Sac*I to insert *su9-nifS<sup>Av</sup>*, and finally digested with *Xho*I to insert *su9-nifM<sup>Av</sup>* by homologous recombination<sup>44</sup>.

All DNA digestions were performed using enzymes from New England Biolabs. Ligated products (T4 ligase, Promega) were introduced into *E. coli* DH5α chemically competent cells and selected on LB (Lysogenic broth) supplemented with appropriate antibiotics. Plasmid extraction was performed using QIaprep Spin Miniprep kit (QIAGEN) and correct cloning was confirmed by Sanger sequencing (Macrogen).

**Growth of *S. cerevisiae*, mitochondria isolations, and ScNifH purifications.** *S. cerevisiae* for galactose-induced expression of ScNifH<sup>Mm</sup>, ScNifH<sup>Mi</sup>, ScNifH<sup>Ht</sup>, and ScNifH<sup>Av</sup> together with SU9-NifU<sup>Av</sup>, SU9-NifS<sup>Av</sup>, and SU9-NifM<sup>Av</sup> (XJ1Y-XJ4Y, Supplementary Table 3) were cultured in 4-l fermenters under aerobic conditions (0.625 l of air per minute and 1 l of culture, 250 rpm stirring) and used for mitochondria isolations or NifH purifications as previously described<sup>11</sup>. Preparation of CFE and STAC purifications were performed at O<sub>2</sub>-levels below 1 ppm in anaerobic chambers (Coy systems or MBraun). Typically, cells were resuspended in lysis buffer (100 mM Tris-HCl (pH 8.6), 200 mM NaCl, 10% glycerol, 2 mM sodium dithionite (DTH), 1 mM PMSF, 1 μg/ml leupeptin, 5 μg/ml DNase I) at a ratio of 1:2 (w/v). Total extracts (TE) were prepared by lysis of the cell suspensions under anaerobic atmosphere using an EmulsiFlex-C5 homogenizer (Avestin Inc.) operating at 20,000 psi. The TE was transferred to centrifuge tubes equipped with sealing closures (Beckman Coulter) and centrifuged at 50,000 g for 1 h at 4 °C (Avanti J-26 XP). The supernatant was filtered using filtering cups with a pore size of 0.2 μm, rendering cell-free extract (CFE) of soluble proteins that was loaded at 2.5 ml/min into a 5 ml Strep-Tactin XP column (IBA LifeSciences) attached to an ÄKTA FPLC (GE Healthcare). The column was washed using 75 ml washing buffer (100 mM Tris-HCl pH 8.0, 200 mM NaCl, 10% glycerol, 2 mM DTH). Strep-Tactin XP column-bound proteins were eluted with 15 ml washing buffer supplemented with 50 mM biotin (IBA LifeSciences). The elution fraction was concentrated, and biotin removed, by passing the protein through PD-10 desalting columns (GE Healthcare). Desalted eluate was further concentrated using centrifugal filters (Amicon, Millipore) with 30 kDa cutoff. Finally, the concentrated protein was snap-frozen in cryovials (Nalgene) and stored in liquid N<sub>2</sub>.

**Soil Fe fertilization, preparation of anaerobic *N. benthamiana* leaf cell-free extracts, and purification of NbNifH<sup>Ht</sup> and NbNifU<sup>Av</sup>.** *N. benthamiana* plants were grown under long day conditions (16 h light/8 h dark) with supporting light from 17:00 to 00:00 for 4 weeks. For Fe fertilization experiments, plants were irrigated (2l per week) with tap water supplemented with 1 g/l Sequestrene G100



(Syngenta). Leaves harvested after extended dark period (16 h) were kept in darkness from 17:00 (previous day) until sample collection (09:00 following morning).

Purifications of *NbNifH<sup>Ht</sup>* and *NbNifU<sup>Av</sup>* were performed at O<sub>2</sub>-levels below 1 ppm inside anaerobic chambers (Coy systema or MBraun). Typically, 200 g of leaf material was harvested and frozen in liquid N<sub>2</sub>. Leaf material was transferred into an anaerobic chamber in frozen condition and disrupted in equal amount (w/v) of lysis buffer (100 mM Tris-HCl pH 8.6, 200 mM NaCl, 10% glycerol, 2 mM DTH, 1 mM PMSF, 1 µg/ml leupeptin, 5 µg/ml DNaseI) using a blender (Oster Classic 4655) operating at maximum power and maintained at 4 °C using a circulating water bath. TE was filtered through cheese cloth to remove larger debris. Preparation CFE by centrifugation, Strep-Tactin affinity chromatography, protein elution, concentration, and storage was identical as for yeast-expressed *ScNifH* proteins. The purification procedure for *NbNifU<sup>Av</sup>* only differed in that no DTH was present in the buffers.

**Protein methods, antibodies, UV-vis absorption spectrum, and electron paramagnetic resonance.** Protein concentrations were measured using the BCA protein assay (PIERCE) in combination with iodoacetamide to eliminate the interfering effect of DTH<sup>45</sup>. Colorimetric Fe determination was performed as reported<sup>46</sup>, and the N-terminal amino acid sequences were determined by Edman degradation (Proteome Factory AG).

Antibodies used in this study and their dilutions for immunoblotting were as follows: polyclonal antibodies detecting *NifU<sup>Av</sup>* (used at 1:2,000 in 5% BSA), *NifS<sup>Av</sup>* (used at 1:1,000 in 5% BSA), *NifH<sup>Av</sup>* (used at 1:5,000 in 5% BSA), *NifM<sup>Av</sup>* (used at 1:2,000 in 5% BSA) were raised against purified preparations of the corresponding *A. vinelandii* proteins (generated *in house*). Strep-tag II ("Strep-MAB", 2-1507-001, IBA Lifesciences, 1:2,000 in 5% BSA), Strep-Tactin conjugated to HRP ("Strep-HRP", 2-1502-001, IBA Lifesciences, 1:50,000 in TBS- T), GFP (sc-9996, Santa Cruz Biotechnology, 1:2,000 in 5% BSA), HSP60 (LK-2, ab59458, Abcam, 1:1,000 in 5% BSA), and Tubulin (3H3087, sc-69971, Santa Cruz Biotechnology, 1:500 in 5% BSA) specific antibodies are commercially available.

The UV-vis absorption spectra were recorded after removal of the DTH from the protein samples using PD-10 desalting columns (GE Healthcare) equilibrated with the corresponding protein buffer without DTH. DTH-free protein samples were then diluted in the same buffers and transferred to a Q6 spectroscopy cuvettes with sealing closures. Absorption (280 nm to 800 nm) was recorded using a UV-2600 spectrophotometer (Shimadzu).

EPR measurements were performed in a Bruker E500 spectrometer equipped with a resonator operating in the TE<sub>102</sub> mode at 9.47 GHz. Temperature was set and stabilized to 10 K by an Oxford temperature controller regulating a gas-flow cryostat refrigerated with helium. For measurements, a microwave power of 2.5 mW and a magnetic field modulation amplitude of 1 mT was used. Experimental conditions were carefully monitored to avoid over-modulation or saturation effects. Simulations of the EPR spectra were performed using the Matlab toolbox Easypin<sup>47</sup>.

**In vitro NifH activity.** NifH activity was determined as described by Shah et al. with slight modifications<sup>48</sup>. Reactions were prepared inside anaerobic chambers. Purified NifH proteins were analyzed by ARA after addition of NifDK<sup>Av</sup> and ATP-regenerating mixture (1.23 mM ATP, 18 mM phosphocreatine, 2.2 mM MgCl<sub>2</sub>, 3 mM DTH and 46 µg/ml of creatine phosphokinase, 22 mM Tris-HCl pH 7.5) in a final volume of 600 µl inside 9 ml serum vials under Ar atmosphere containing 500 µl of acetylene (1 atm). The ratio of NifH to NifDK in the assays was 40:1 unless otherwise indicated. The ARA were performed at 30 °C in a shaking water bath for 15 min. Reactions were stopped by adding 100 µl of 8 M NaOH. Positive control reactions for acetylene reduction were carried out with NifH<sup>Av</sup>. Ethylene formed was measured in 50 µl gas phase samples using a Porapak N 80/100 column in a gas chromatograph (Shimadzu).

Reduction of N<sub>2</sub> to NH<sub>3</sub> was determined in reaction mixtures prepared as for the ARA but containing 100 mM 3-(N-morpholino)propanesulfonic acid (MOPS), pH 7.8, as buffer. Mixtures were prepared in volumes of 750 µl, from which 100 µl was removed at assay start to serve as background (*t*<sub>0</sub>) for NH<sub>3</sub> measurements. After exchanging vial atmosphere for N<sub>2</sub>, mixtures were incubated at 30 °C for 30 min, and reactions were stopped by addition of 100 µl 5 M EDTA. Twenty-five µl of the blank (*t*<sub>0</sub>) and the reaction (*t*<sub>30</sub>) were added in duplicates to 200 µl o-phthalaldehyde reagent solution (ThermoFisher Scientific) in 96-well microplate for fluorescence-based assays (Nunc). Fluorescence (Ex 390 nm, Em 472 nm) was measured using a Varioskan LUX plate reader (ThermoFisher Scientific). NH<sub>3</sub> production was determined from the increase in fluorescence (*t*<sub>30</sub>-*t*<sub>0</sub>) against standards prepared with NH<sub>4</sub>Cl and recorded in the same plate.

**In vitro P-cluster maturation.** P-cluster maturation assays were performed inside anaerobic chambers. The *in vitro* assay combined isolated NifH to be tested (50 µg) with *A. vinelandii* DJ77 (*ΔnifH*) CFE (4.34 mg total protein) and an excess of pure FeMo-co (0.85 µM) in 500 µl ATP-regenerating mixture as described above. Reactions were incubated at 30 °C for 30 min. Forty µl of 1 mM (NH<sub>4</sub>)<sub>2</sub>MoS<sub>4</sub> (tetrathiomolybdate) were then added and mixtures were incubated for 10 min at

room temperature to prevent further FeMo-co incorporation into NifDK<sup>Av</sup> during the ARA.

Apo-NifDK<sup>Av</sup> activation after P-cluster maturation and FeMo-co insertion was analyzed by ARA after addition of an excess of the same NifH species (100 µg) and ATP-regenerating mixture in a final reaction volume of 1 ml. ARA was carried out in 9 ml serum vials containing Ar and 500 µl of acetylene (1 atm) in the headspace for 15 min at 30 °C. Positive control reactions for *in vitro* P-cluster maturation and ARA contained purified NifH<sup>Av</sup>. Ethylene formed was measured in 50 µl gas phase samples using a Porapak N 80/100 column in a gas chromatograph (Shimadzu).

**In vitro FeMo-co synthesis and apo-NifDK<sup>Av</sup> reconstitution.** NifB-co-dependent FeMo-co synthesis assays were performed inside anaerobic chambers as described by Curatti et al., with slight modifications<sup>34</sup>. One hundred µl reactions contained 3.0 µM NifH, GST-NifX-NifB-co (20.4 µM Fe), 1.5 µM apo-NifEN<sup>Av</sup>, 0.6 µM apo-NifDK<sup>Av</sup>, 17.5 µM Na<sub>2</sub>MoO<sub>4</sub>, 175 µM *R*-homocitrate, 1 mg/ml BSA, and ATP-regenerating mixture (1.23 mM ATP, 18 mM phosphocreatine disodium salt, 2.2 mM MgCl<sub>2</sub>, 3 mM DTH, 46 µg/ml creatine phosphokinase, final concentrations in 22 mM Tris-HCl (pH 7.5) buffer at 30 °C for 60 min.

NifB-dependent FeMo-co synthesis assays were performed as the above described NifB-co-dependent assay replacing GST-NifX-NifB-co by 10.0 µM NifB monomer, 125 µM FeSO<sub>4</sub>, 125 µM Na<sub>2</sub>S, and 125 µM SAM.

Following *in vitro* synthesis of FeMo-co, 17.5 µM (NH<sub>4</sub>)<sub>2</sub>MoS<sub>4</sub> was added to prevent further FeMo-co incorporation into apo-NifDK<sup>Av</sup>, and incubated for 10 min at 25 °C. Activation of apo-NifDK<sup>Av</sup> was analyzed by addition of 500 µl ATP-regenerating mixture and *ScNifH<sup>Ht</sup>* (2.0 µM final concentration) in 9 ml vials containing Ar and 500 µl acetylene. The ARA were performed at 30 °C for 20 min. Positive control reactions for ARA contained NifDK<sup>Av</sup> and NifH<sup>Av</sup>. Ethylene formed was measured in 50 µl gas phase samples using a Porapak N 80/100 column in a gas chromatograph (Shimadzu).

**In vitro [Fe-S] cluster reconstitution and NifH activity.** *In vitro* [Fe-S] cluster reconstitutions of NifH and NifU<sup>Av</sup> purified from *E. coli* (*EcNifU<sup>Av</sup>*)<sup>10</sup> were performed in anaerobic chambers as described by Zheng and Dean<sup>49</sup> with slight modifications. NifH or NifU (20 µM) was added to 22 mM Tris-HCl (pH 7.5) buffer supplemented with 8 mM 1,4-dithiothreitol (DTT) in a final volume of 100 µl and incubated at 37 °C for 30 min. Then, reactions were supplemented with 1 mM L-cysteine, 1 mM DTT, 400 µM (NH<sub>4</sub>)<sub>2</sub>Fe(SO<sub>4</sub>)<sub>2</sub>, and 225 nM NifS<sup>Av</sup> purified from *E. coli* (*EcNifS<sup>Av</sup>*)<sup>10</sup>, and incubated at 37 °C overnight. Finally, the proteins were diluted 1000-fold in 22 mM Tris-HCl (pH 7.5) buffer, and then concentrated using centrifugal filters (Amicon, Millipore) with 30 kDa cutoff to remove excess reagents.

For "direct reconstitution" activity assays, the activity of [Fe<sub>4</sub>S<sub>4</sub>] cluster reconstituted NifH protein was determined using ARA. For "NifU-mediated reconstitution", as-isolated NifH protein was mixed with [Fe-S] cluster reconstituted *EcNifU<sup>Av</sup>*, and then immediately used for ARA.

**Statistics and reproducibility.** Distinct samples were used for *in vitro* activity measurements and sample sizes are indicated by *n*, where each distinct sample was measured at least two times. Mean of measured activities are shown. The data presented in the figure graphs are listed in Supplementary Data 2.

## Data availability

The authors declare that the data supporting the findings of this study are available within the article, its supplementary information and data, and upon request.

Received: 15 June 2020; Accepted: 25 November 2020;  
Published online: 04 January 2021

## References

1. Erisman, J. W. et al. *Nitrogen: Too much of a vital resource*. (Science Brief. WWF Netherlands, Zeist, The Netherlands, 2015).
2. Borlaug, N. E. Feeding a world of 10 billion people: the miracle ahead. *In Vitro Cell Dev. Biol.* **38**, 221–228 (2002).
3. Sibhatu, K. T. & Qaim, M. Rural food security, subsistence agriculture, and seasonality. *PLoS One* **12**, e0186406 (2017).
4. Bloch, S. E., Ryu, M. H., Ozaydin, B. & Broglie, R. Harnessing atmospheric nitrogen for cereal crop production. *Curr. Opin. Biotechnol.* **62**, 181–188 (2020).
5. Burén, S. & Rubio, L. M. State of the art in eukaryotic nitrogenase engineering. *FEMS Microbiol. Lett.* **365**, fnx274–fnx274 (2018).
6. Bulen, W. A. & LeComte, J. R. The nitrogenase system from *Azotobacter*: two-enzyme requirement for N<sub>2</sub> reduction, ATP-dependent H<sub>2</sub> evolution, and ATP hydrolysis. *Proc. Natl Acad. Sci. USA* **56**, 979–986 (1966).

7. Hoffman, B. M., Lukoyanov, D., Yang, Z. Y., Dean, D. R. & Seefeldt, L. C. Mechanism of nitrogen fixation by nitrogenase: the next stage. *Chem. Rev.* **114**, 4041–4062 (2014).
8. Seefeldt, L. C. et al. Reduction of substrates by nitrogenases. *Chem. Rev.* **120**, 5082–5106 (2020).
9. Burén, S., Jimenez-Vicente, E., Echavarrri-Erasun, C. & Rubio, L. M. Biosynthesis of nitrogenase cofactors. *Chem. Rev.* **120**, 4921–4968 (2020).
10. Lopez-Torrejon, G. et al. Expression of a functional oxygen-labile nitrogenase component in the mitochondrial matrix of aerobically grown yeast. *Nat. Commun.* **7**, 11426 (2016).
11. Burén, S. et al. Biosynthesis of the nitrogenase active-site cofactor precursor NiFB-co in *Saccharomyces cerevisiae*. *Proc. Natl Acad. Sci. USA* **116**, 25078–25086 (2019).
12. Eserverri, A. et al. Use of synthetic biology tools to optimize the production of active nitrogenase Fe protein in chloroplasts of tobacco leaf cells. *Plant Biotechnol. J.* **18**, 1882–1896 (2020).
13. Burén, S. et al. Formation of nitrogenase NifDK tetramers in the mitochondria of *Saccharomyces cerevisiae*. *ACS Synth. Biol.* **6**, 1043–1055 (2017).
14. Allen, R. S. et al. Plant expression of NifD protein variants resistant to mitochondrial degradation. *Proc. Natl Acad. Sci. USA* **117**, 23165–23173 (2020).
15. Xiang, N. et al. Using synthetic biology to overcome barriers to stable expression of nitrogenase in eukaryotic organelles. *Proc. Natl Acad. Sci. USA* **117**, 16537–16545 (2020).
16. Okada, S. et al. A Synthetic Biology Workflow Reveals Variation in Processing and Solubility of Nitrogenase Proteins Targeted to Plant Mitochondria, and Differing Tolerance of Targeting Sequences in a Bacterial Nitrogenase Assay. *Front Plant Sci* **11**, 552160 (2020).
17. Poza-Carrion, C., Jimenez-Vicente, E., Navarro-Rodriguez, M., Echavarrri-Erasun, C. & Rubio, L. M. Kinetics of Nif gene expression in a nitrogen-fixing bacterium. *J. Bacteriol.* **196**, 595–603 (2014).
18. Vogtle, F. N. et al. Global analysis of the mitochondrial N-proteome identifies a processing peptidase critical for protein stability. *Cell* **139**, 428–439 (2009).
19. Pollegioni, L., Schonbrunn, E. & Siehl, D. Molecular basis of glyphosate resistance-different approaches through protein engineering. *FEBS J.* **278**, 2753–2766 (2011).
20. Paine, J. A. et al. Improving the nutritional value of golden rice through increased pro-vitamin A content. *Nat. Biotechnol.* **23**, 482–487 (2005).
21. Nakamura, Y., Gojobori, T. & Ikemura, T. Codon usage tabulated from international DNA sequence databases: status for the year 2000. *Nucleic Acids Res.* **28**, 292 (2000).
22. Schmidt, T. G. et al. Development of the Twin-Strep-tag(R) and its application for purification of recombinant proteins from cell culture supernatants. *Protein Expr. Purif.* **92**, 54–61 (2013).
23. Westermann, B. & Neupert, W. Mitochondria-targeted green fluorescent proteins: convenient tools for the study of organelle biogenesis in *Saccharomyces cerevisiae*. *Yeast* **16**, 1421–1427 (2000).
24. Burén, S., Jiang, X., Lopez-Torrejon, G., Echavarrri-Erasun, C. & Rubio, L. M. Purification and in vitro activity of mitochondria targeted nitrogenase cofactor maturase NiFB. *Front. Plant Sci.* **8**, 1567 (2017).
25. Kennedy, C. & Dean, D. The nifU, nifS and nifV gene products are required for activity of all three nitrogenases of *Azotobacter vinelandii*. *Mol. Gen. Genet.* **231**, 494–498 (1992).
26. Howard, K. S. et al. *Klebsiella pneumoniae* nifM gene product is required for stabilization and activation of nitrogenase iron protein in *Escherichia coli*. *J. Biol. Chem.* **261**, 772–778 (1986).
27. Naim, F. et al. Advanced engineering of lipid metabolism in *Nicotiana benthamiana* using a draft genome and the V2 viral silencing-suppressor protein. *PLoS One* **7**, e52717 (2012).
28. Jimenez-Vicente, E. et al. Sequential and differential interaction of assembly factors during nitrogenase MoFe protein maturation. *J. Biol. Chem.* **293**, 9812–9823 (2018).
29. Baker, R. T. & Varshavsky, A. Yeast N-terminal amidase. A new enzyme and component of the N-end rule pathway. *J. Biol. Chem.* **270**, 12065–12074 (1995).
30. Jacobson, M. R. et al. Physical and genetic map of the major nif gene cluster from *Azotobacter vinelandii*. *J. Bacteriol.* **171**, 1017–1027 (1989).
31. Christiansen, J., Goodwin, P. J., Lanzilotta, W. N., Seefeldt, L. C. & Dean, D. R. Catalytic and biophysical properties of a nitrogenase Apo-MoFe protein produced by a nifB-deletion mutant of *Azotobacter vinelandii*. *Biochemistry* **37**, 12611–12623 (1998).
32. Hernandez, J. A. et al. NifX and NifEN exchange NiFB cofactor and the VK-cluster, a newly isolated intermediate of the iron-molybdenum cofactor biosynthetic pathway. *Mol. Microbiol.* **63**, 177–192 (2007).
33. Guo, Y. et al. The nitrogenase FeMo-cofactor precursor formed by NiFB protein: a diamagnetic cluster containing eight iron atoms. *Angewandte Chem. Int. Ed. Engl.* **55**, 12764–12767 (2016).
34. Curatti, L. et al. In vitro synthesis of the iron-molybdenum cofactor of nitrogenase from iron, sulfur, molybdenum, and homocitrate using purified proteins. *Proc. Natl Acad. Sci. USA* **104**, 17626–17631 (2007).
35. Emerich, D. W. & Burris, R. H. Complementary functioning of the component proteins of nitrogenase from several bacteria. *J. Bacteriol.* **134**, 936–943 (1978).
36. Fu, W., Jack, R. F., Morgan, T. V., Dean, D. R. & Johnson, M. K. nifU gene product from *Azotobacter vinelandii* is a homodimer that contains two identical [2Fe-2S] clusters. *Biochemistry* **33**, 13455–13463 (1994).
37. Koonin, E. V. Origin of eukaryotes from within archaea, archaeal eukaryome and bursts of gene gain: eukaryogenesis just made easier? *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **370**, 20140333 (2015).
38. Temme, K., Zhao, D. & Voigt, C. A. Refactoring the nitrogen fixation gene cluster from *Klebsiella oxytoca*. *Proc. Natl Acad. Sci. USA* **109**, 7085–7090 (2012).
39. Yang, J. et al. Polyprotein strategy for stoichiometric assembly of nitrogen fixation components for synthetic biology. *Proc. Natl Acad. Sci. USA* **115**, E8509–E8517 (2018).
40. Chrétien, D. et al. Mitochondria are physiologically maintained at close to 50 °C. *PLOS Biol.* **16**, e2003992–e2003992 (2018).
41. Arragain, S. et al. Diversity and functional analysis of the FeMo-cofactor maturase NiFB. *Front. Plant Sci.* **8**, 1947 (2017).
42. Nishihara, A. et al. Nitrogenase activity in thermophilic chemolithoautotrophic bacteria in the phylum aquificae isolated under nitrogen-fixing conditions from Nakabusa hot springs. *Microbes Environ.* **33**, 394–401 (2018).
43. Zeytun, A. et al. Complete genome sequence of *Hydrogenobacter thermophilus* type strain (TK-6). *Stand. Genom. Sci.* **4**, 131–143 (2011).
44. Koskela, E. V. & Frey, A. D. Homologous recombinatorial cloning without the creation of single-stranded ends: exonuclease and ligation-independent cloning (ELIC). *Mol. Biotechnol.* **57**, 233–240 (2015).
45. Hill, H. D. & Straka, J. G. Protein determination using bicinchoninic acid in the presence of sulfhydryl reagents. *Anal. Biochem.* **170**, 203–208 (1988).
46. Fish, W. W. Rapid colorimetric micromethod for the quantitation of complexed iron in biological samples. *Methods Enzymol.* **158**, 357–364 (1988).
47. Stoll, S. & Schweiger, A. EasySpin, a comprehensive software package for spectral simulation and analysis in EPR. *J. Magn. Reson.* **178**, 42–55 (2006).
48. Shah, V. K. & Brill, W. J. Nitrogenase. IV. Simple method of purification to homogeneity of nitrogenase components from *Azotobacter vinelandii*. *Biochim. Biophys. Acta* **305**, 445–454 (1973).
49. Zheng, L. & Dean, D. R. Catalytic formation of a nitrogenase iron-sulfur cluster. *J. Biol. Chem.* **269**, 18723–18726 (1994).

## Acknowledgements

This paper is dedicated to the memory of Prof. Tomás Ruiz Argüeso. We thank Marcel Veldhuizen for yeast fermentations and Carlos Echavarrri-Erasun for helpful discussions. Funding for this research was provided by Bill & Melinda Gates Foundation Grant OPP1143172 (L.M.R.). X.J. is supported by a doctoral fellowship from Universidad Politécnica de Madrid, and L.P.T. is recipient of the FPU16/02284 from Ministerio de Ciencia, Innovación y Universidades. I.G.R. acknowledges financial contributions from MINECO (CTQ2015-64486-R), Gobierno de Aragón (E35\_17R) and Fondo Social Europeo “Construyendo Europa desde Aragón”.

## Author contributions

X.J., L.P.T., D.C., I.G.R., R.C.R., Á.E., G.L.T., and S.B. performed the experiments; X.J., L.P.T., I.G.R., S.B., and L.M.R. designed experiments and analyzed data; X.J., S.B., and L.M.R. wrote the paper.

## Competing interests

The authors declare no competing interests.

## Additional information

Supplementary information is available for this paper at <https://doi.org/10.1038/s42003-020-01536-6>.

Correspondence and requests for materials should be addressed to S.Bén. or L.M.R.

Reprints and permission information is available at <http://www.nature.com/reprints>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2021